A Feasibility Study Based on the Framingham Data

March 12, 2015

Abstract

Given the evidence of peer pressure impacting human decision-making and the fact that growing online social networks now make us more connected than ever, further advances are needed and become possible to enable accurate causal analyses in the presence of network effects. Separating the challenges of *dependent self-selection* and *interference* in network settings, once can distinguish three groups of program evaluation problems: those with dependent self-selection but no interference, those with interference but no dependent self-selection, and those with both. This paper outlines a new research agenda for quantifying the effects of social network dependencies on self-selection by advancing the theories and applications of two research areas - Exponential Random Graph Modeling and "propensity score" based observational causal inference.

The paper then reports on the design and results of a feasibility study where the expanded propensity score-based causal inference method is applied to the Framingham Heart Study dataset of socially connected individuals; the query of interest is the causal effect of smoking on BMI (weight indicator). The paper covers the steps taken in data pre-processing, social network clustering, ERGM parameter estimation experiences, evaluation of expanded propensity scores, and estimation of average treatment effects. In this investigation, it is established that new stable-network ERGMs can successfully inform the expanded propensity score estimation. Moreover, causal analyses with and without blocking the network effects are found to produce different causal estimates, indicating that the network structure is indeed a confounding factor in causal analyses in connected communities.

1 Introduction and Motivation

The paper presents a new research agenda towards enabling program effectiveness evaluation based on observations from a social network of human subjects, self-selecting themselves for treatment. The research explores the impact that social network effects may have on voluntary participation of connected units in a community-wide program. Understanding the network-induced dependent self-selection could help distill peer influence mechanisms, accurately evaluate program casual effects and inform planning of program dissemination activities.

Under the premise that connected units make/revise their individual decisions, reacting to the influence form their peers, the paper explores how exponential random graph modeling and propensity score estimation can be combined to advance program evaluation research in network settings. We present the new approach that incorporates network-based measures into propensity score theory, accommodating, and under reasonable assumptions, separating network effects from individual effects in explaining units' program participation decisions.

1.1 Background: Program Evaluation

The problem of program/policy evaluation has received much attention from theoreticians and applied scientists. A broadly stated program evaluation question is: "What is the effect of a program on participants and nonparticipants compared to no program at all or some alternative program?" An excellent in-depth coverage of decision problems in policy analysis is offered in the Noble Prize lecture by Heckman (2000).

Multiple research domains have addressed the facets of the program evaluation challenge. One such domain, originating from the works of Rubin and Rosenbaum, focuses on making inference from limited data obtained by observation (Rubin, 1973, 1977, 1979, 1980; Rosenbaum and Rubin, 1983; Rosenbaum, 1989, 1991). Observational data are more often available to researchers compared to data collected in randomized experiments. Numerous research efforts relied on observational data in micro econometrics (LaLonde, 1986; Heckman et al., 1998, 1999; Heckman and Vytlacil, 2001), political science (Herron and Wand, 2007; Imai, 2005; Sekhon, 2008), medicine (Rubin, 1991; da Veiga and Wilder, 2008), and psychology (Reinisch et al., 1995), among others.

Selection bias is the root of the challenge of evaluating causal effect, i.e., the effect of a program isolated from that of confounding factors (Heckman, 2000). Such confounding factors are typically represented by the covariates of units under study that by default differentiate the units in the entire population, and hence, the outcomes measured on them. Selection bias arises when (potential) individual outcomes over the underlying population of interest are only partially observed or whenever a rule other than simple random sampling is used to sample the population.

Self-selection bias arises in the estimation of causal effects of programs/treatments from limited observed data, when study participants select themselves for treatment. Self-selection makes it difficult to establish causation, and complicates research about programs or products (Jacobs et al., 2009). The challenge of non-random sampling is exacerbated when evaluating causal effects in networked communities, in the presence of explicit dyadic dependencies between individuals. For example, in enrolling into an extracurricular program at school, students' subjective views of the program and the decisions to participate in it may in large part depend on the respective views/decisions of their friends (in addition to individual preferences that can be explained by their covariates).

1.2 Background: Peer Effects and Social Network Analysis

A growing body of literature on peer effects aims at measuring the induced effect of connection on the individual behavior and decisions (Jackson, 2010; Aral et al., 2009; Zimmerman, 2003). These types of dependencies that may define "peers" include friendly relationships, roommates, classmates, colleagues, neighbor etc. One class of peer effect studies encompasses mathematical models of long run behaviors/equilibria of social interactions of peer groups (Jackson, 2010). The other focuses on the identification and estimation of peer effects. Among such investigations, peer pressure in adolescent communities has long been an important topic of research (Bauman and Ennett, 1996; Fletcher et al., 1995). Studies of homophily and contagion effects on health-related decisions are most notable among them (Fowler and Christakis, 2010; Christakis and Fowler, 2008, 2007).

Researchers have begun to approach the problem of estimating casual effects (Oakes, 2004) of the peer influence (also known as neighborhood effects or social capital) in the 1990 (Heckman et al., 1998). Many agree that randomized control trials, in which peer effect can be blocked, are practically difficult to stage; these reasons are well illustrated by the statement of Jencks and

Mayer (1990) who studied segregation and the social environment in poor neighborhoods: "... the best way to estimate neighborhood effects would be to conduct controlled experiments in which we assigned families randomly to different neighborhoods, persuaded each family to remain in its assigned neighborhood for a protracted period, and then measured each neighborhood's effect on the children involved". Recent studies (Sampson, 1968) managed to assess the statistical significance of peer effects by pure observation (Jaccard et al., 2005; Mouw, 2006) and by large-scale experiments (Aral and Walker, 2011; Carrell et al., 2008). In all these cases, network effects have been found significant in explaining participants' decisions; however, no-question-based method relating peer effects with underlying network structure yet exists. Note that most peer effects investigations considered problems where exist dependency structures between the studied units were unknown (Heckman et al., 1998), while many other faced problems where networks were themselves perturbed by evaluated phenomena (Aral et al., 2009). Such complex settings make the task of isolating network effects especially difficult to approach.

The branch of literature on peer effects have been developing in the parallel with the advances in social network analysis (SNA) (Scott and Carrington, 2011; Zimmerman, 2003). The SNA communities use graphs as a convenient way of representing multidimensional relational data. Nodes and edges of a graph can contain information in the form of attributes. Then, the graph is called attributed. Central of the SNA research is the notion of local neighborhood. It is motivated by the notion that the decisions/outcomes of an individual directly depend only on the decision/outcomes of such other individuals that are close to them. Exact definition of a local neighborhood may vary across problem. It is typically defined based on elementary graph-based structures that represent pairwise and higher-order network dependencies, akin to the model terms in regression analysis. SNA metrics, models and methods have designed for the study of properties of graphs and making inference about connected social units (Freeman, 1979; Hinde, 1976; Bonacich, 1972; Luce and Perry, 1949; Handcock et al., 2003; Snijders et al., 2010; Butts, 2008; Leskovec et al., 2010). Much attention in this work has been paid to describing network formation, i.e. creating models of networks (Watts and Strogatz, 1998; Albert and Barabási, 2002; Robins et al., 2007). The most notable effects emphasizing individual differences between network actors employed exponential random graph models of network formation and longitudinal actor-oriented and event oriented model (Snijders et al., 2010; Butts, 2008).

A gap between program evaluation and peer effects investigations, which this research seeks to fill, is that of explaining behavior on networks. In particular, it would be beneficial to identify the effects of the network structure on decision-making under the assumption that the studied units are embedded in a known stable network, the decision of the units participating in an evaluated program can be viewed as self-assigned treatments and studied within the conventional framework of potential outcomes(Rubin, 2005; Lindquist and Sobel, 2011), used in observational casual inference.

1.3 Background and Terminology: Casual Inference with Observational Data

For unit u, let $Y_u^1(Y_u^0)$ denote a treated (untreated) outcome, T_u a treatment/participation indicator (1 means treated, 0 means not treated), and $X_u = \{X_{1u}, X_{2u}, ..., X_{Ku}\}$ a vector of values for K covariates. In both experimental and observational settings, a population of units is under consideration. For a particular unit u, the causal effect of the treatment (relative to the control) is defined as the difference in outcomes that would result from receiving and not receiving the treatment, $Y_u^1 - Y_u^0$. The fundamental problem of causal inference is that it is impossible to observe both values Y_u^1 and Y_u^0 on the same unit u (Holland, 1986) (e.g., a person either smokes or does not

smoke). The outcome of an observation of a unit is termed the observed response, $T_u Y_u^1 + (1-T_u)Y_u^0$. The Rubin causal model (Rubin, 1974, 1978) conceptualizes this causal inference framework so that the response under either treatment or control, but not both, needs to be observed for each unit. That is, one statistical solution to the fundamental problem of causal inference is to shift to an examination of an average causal effect over all units in the population, $E(Y_u^1 - Y_u^0) = E(Y_u^1) - E(Y_u^0)$, where $E(Y_u^1)$ is computed from the treatment group and $E(Y_u^0)$ is computed from the control group.

An important consideration is how one determines which units will inform the values of Y_u^1 and Y_u^0. In an observational study, one observes some pool of units who have received a treatment, giving $E(Y_u^1|T=1)$, and some pool of units who have not received this treatment, giving $E(Y_u^0|T=0)$. In general, $E(Y_u^1) \neq E(Y_u^1|T=1)$ and $E(Y_u^0) \neq E(Y_u^0|T=0)$. The two treatment effects that are typically used in practical studies are the average treatment effect (ATE), $E(Y_u^1 - Y_u^0)$, and average treatment effect for the treated (ATT), $E(Y_u^1|T=1) - E(Y_u^1|T=1)$. If exposure to treatment ($T=1$) or control ($T=0$) is statistically independent of response and covariate values, then the units have been properly randomized into treatment and control pools, rendering ATE and ATT to be the same. This situation is not typically the case in observational studies since units are not randomly placed into treatment and control pools. Instead $ATT = E(Y_u^1|T=1) - E(Y_u^1|T=1) = E(Y_u^1|T=1) - E(Y_u^1|T=0) + \beta$, where self-selection bias is present, defined as $\beta \equiv E(Y_u^0|T=0) - E(Y_u^0|T=0)$.

Confounding effect contributing to the bias may exist in a dataset due to both observed (those reflected in the dataset) and unobserved covariates. Dealing with unobservable covariates is a fundamental challenge for causal inference and requires additional information to supplement the available data, whereas the effects of observed covariates can be isolated by data post-processing.

Among conventional methods for observational causal inference, the most notable is based on the idea of matching. Matching is a procedure that involves dropping/repeating/grouping observations from a given dataset to ensure that the empirical distributions of the observed covariates over the compared treatment/control (participant/nonparticipant) groups are identical, in which case the two groups are called "balanced" (Imbens, 2000; Rubin, 2006; Abadie and Imbens, 2006). As such, the non-random nature of self-selection forces program evaluation to rely only on matching, implemented after the participation decisions have been made and serves as a tool for reducing bias due to observable covariates.

The most widely-used matching methods are Mahalanobis distance based matching (Rubin, 1980) and propensity score (PS) based matching (Rosenbaum and Rubin, 1983). Mahalanobis distance matching directly matchs covariates of individuals. However, PS based matching does that indirectly. It is found that PS is more robust than Mahalanobis distance based matching. Propensity score is defined as the probability of an individual in a population to participate in the program, or in other words, be self-selected to receive treatment. Given a dataset, propensity scores $P(T_u = 1|X_u = X_u^{'})$ of units (u) in the underlying population can be estimated via regression with the observed treatments and covariates. Having selected a suitable regression model, one can use the resulting propensity scores as the weights associated with observed unit responses to estimate ATE or conduct matching of the propensity scores to estimate ATT (Rosenbaum and Rubin, 1983; Abadie and Imbens, 2002), and investigate the properties of estimator variances (Abadie, 2002; Abadie and Imbens, 2006; Caliendo and Kopeinig, 2008; Tu and Zhou, 2002).

While the propensity score based approach remains the primary inference tool for the majority of observational scientists, some alternative methodologies have been recently proposed. Recently,

Balance Optimization Subset selection approach has been developed for causal inference with observational data relying on explicit balance optimization (Nikolaev et al., 2013; Cho et al., 2011) [106]. The approach re-casts the problem of matching individual units into a subset selection problem that can be tackled by specially-designed discrete optimization algorithms on an aggregate, group level (Abadie and Imbens, 2006; Abadie and Gardeazabal, 2003). Other recent work on improving matching procedure included attempts to streamline the process of match - check balance adjust and repeat as needed by using a genetic algorithm (Diamond and Sekhon, 2013; Sekhon, 2008) and the introduction of the notion of fine balance, which refers to exactly balancing a nominal variable, often one with many categories, without trying to match individuals on this variable (Rosenbaum et al., 2007).

Conventional methods for observational causal inference cannot be directly applied to study connected communities. Peer effects literature reports that in network settings, one key assumption commonly made within the potential outcomes framework does not hold: it is the assumption of conditional ignorability, stating that conditional on covariates, potential outcomes are independent of treatments. However, the fact that social networks have well-defined and known dyadic structure of units dependencies might help enable accurate inference.

1.4 Network Effects in Program Evaluations: Challenges and Opportunities

The network context creates new challenges for program evaluation research. It implies that both program participation and outcomes in a population may depend, in a structured way (through dyadic relationships), not only on the real-valued covariates of individuals but also on the neighborhoods in which in the individuals are embedded. Both the structure of connection to and between their neighbors and those neighbors' covariates, program participation and outcomes.

Social interaction is a major source of bias in studies with connected human subjects (Holland, 1986). Two distinct types of peer effects, relevant in the program evaluation context, can be distinguished: one is dependent self-selection and the other is inference. Using the causal inference terminology, dependent self-selection implies dependent treatment assignments. In turn, inference implies treating some units may affect the outcomes of other units (Sobel, 2006; Hudgens and Halloran, 2008). For example, vaccine given to an individual may prevent others around them from being infected, by hampering the decrease transmission over the contact network (Halloran and Struchiner, 1995; David and Kempton, 1996). Under interference, the stable Unit Treatment value assumption, fundamental for the potential outcomes framework is violated (Cox, 1958; Rubin, 1978, 1980).

Program evaluation problem instances on social networks can be split into three groups: those with dependent self-selection but interference, those with interference but no dependent self-selection, and those with both. Each problem instance must be treated individually to determine which group it belongs to. Consider the following illustrative examples.
Example 1: A weight loss program is introduced to a workplace community. Some community sign up, and in two methods. Their outcomes are compared against the rest of the community to assess the program's effectiveness.
Example 2: A farmer uses an insecticide on several plots of his land (but not all). After the season, he looks to evaluate the insecticide by comparing the yield from the treated vs. untreated plots.
Example 3: An extracurricular SAT course is offered to the seniors in a local high school. The students share opinions about the program prior to enrolling, and some eventually enroll. A year

later, policy maker contemplates about expanding the program, looks to evaluate programs effect on its participants, based on the observed SAT scores of the students at the test school..

In Example 1, one has to deal with dependent self-selection, whereas interference is unlikely: adults watching their weight are known to tend to go to a gym and maintain diet on an individual basis. In Example 2, the treatments are given (there is no self-selection); however, interference arises, since applying the insecticide to one plot may reduce the number of insects on the adjacent plots. In Example 3, dependent self-selection and interference are both present: students are likely to both persuade each other to take or skip the course in groups, and also, study together.

It is undesirable to begin building program evaluation theory on networks by simultaneously addressing dependent self-selection and interference challenges, because of the difficulty in separating their induced influences. As explained above, dependent self-selection and interference violate assumptions of different meanings and magnitudes, and hence the effect of each can be better interpreted in isolation from each other. Meanwhile, focusing the attention exclusively on self-selection on networks serves a well-defined purpose: understanding peer pressure impact with respect to a specific causal query/program: e.g. a friendship network can impose different influences on the same high school students as they make decisions about choosing a mobile brand vs enrolling into an after hours program/class. Quantification of peer pressure, captured by network structure, would help address the questions of peer pressure, captured by network structure, would help address the questions:

(1) How importance the observations of network structure are for human subjects research?

(2) What elements of this structure are critical for maintaining peer pressure with respect to a given problem?

(3) What measures are the best suited for evaluating a total impact of a program on a connected community?

(4) How the results of a program, studied on one network, can be extrapolated to other networks?

(5) What dissemination strategy is the most effective for a program?

2 Research Objectives

The objective of this research is to enable accurate program evaluation based on observations from a social network of human subjects, by understanding the impacts of network-based self-selection on program participant pool and consequently, program effect on the population.

The idea of quantifying network effects and removing their bias in causal investigations for programs with voluntary participation is the central to this research. In order to quantify the relationship between the network structure and program evaluation. It is assumed that the program exposed population forms a stable connection structure, which in not influenced by the decisions of non-participant decisions. The described setup is justified in various circumstances:

(1) people enrolling into educational programs

(2) buying "Groupon"

(3) responding to online surveys and research participation advertisements

(4) following online blogs

(5) installing energy-saving appliances

(6) turning out for elections, etc.

Observe that the program outcomes in these described settings are individual-based: once one decides to participate in such programs, the resulting intellectual skills or electricity savings in one's individual gain; A groupon holder applies the discount only to their own dinner; a person casts his/her vote independently of others showing up at the booth; Once enrolled, one will respond to paid survey with their own opinions etc. Observe also that, while people learn about various programs from friends and family members, and make participation decisions collectively, they do not tend to make or lose friends over participation in most programs. Relying on this logic, a special attention is paid to studying network influence on decision-making in the settings where dependent self-selection is present and interference is not.

By identifying network-based confounding factors, we propose expansion of the concept of propensity score which depends on individual covariate and network covariate. In order to extend the propensity score based analytical toolbox to be used in social network settings. We develop an approach to estimate propensity scores from conditional distribution of treatment given individual covariate and social network. We called it "expanded propensity score". Using the expanded propensity score, we have designed estimators newly defined program/policy effect measures on social networks. To summarize, we provided a foundation for designing program effect estimation theory based on observational network data, and creates a methodological toolbox for the benefit of the scientific communities engaged in studies for peer effects and program/policy evaluation planning.

3 Distilling Network Effect Ingredients

Dependencies between social network actors have been examined in studies of network formation and peer effects (Jackson, 2010; Aral et al., 2009; Zimmerman, 2003). The roots of collective processes lie in in sociology/communication theories, competing against and reinforcing each other. Cognitive theories, contagion theories, theories of (mutual) self-interest, homophiles, social balance rely on the mechanisms that can be expressed by such measures as reciprocation, transitivity, cyclicality, centralization and so on (Monge and Contractor, 2003; Contractor et al., 2006). These measures depend on observable patterns of network ties and actors' individual covariates. Each distinct pattern can be decomposed into and quantified by the count of network ingredients, known as "Network statistics". This approach has been successful in network formation modeling (Snijders et al., 2010; Robins et al., 2007); we proposed to extend its use to quantify decision-making on networks. To the end, special network statistics can be used to expresses the dependencies between program participation decisions of connected units, their covariates and the network statistics.

Fig. 1 depicts a friendship network of boys and girls responding to an extracurricular school program. In order to explain the dependencies in the students' self-selection into the program, one can use network statistics. The count of the first network statistics in the Fig. 1 informs one about how much friends tend to participate in the program together, independent of the gender. The second count quantifies how often boyfriends make some participation decision. The third one indicates if boys tend to conform to the decision agreed upon by two of their friends. The vector of counts for these and other network statistics in a given network distinguishes the observed mechanisms.

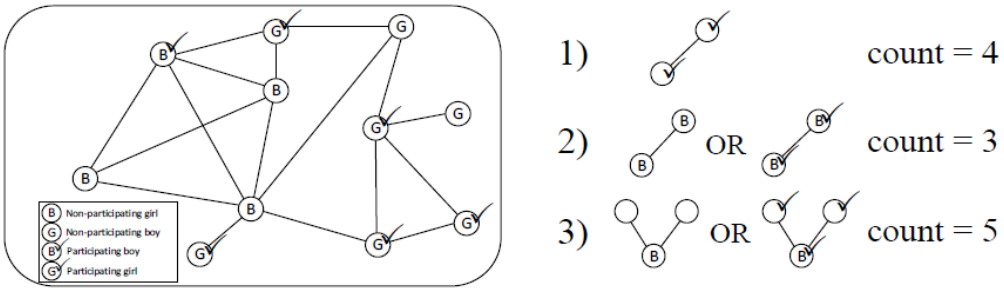


Figure 1: Example Network statistics for representing participation decision dependence

One can also investigate which network statistics are significant in exploring the variation in participation. Given a unit with a known covariate vector, embedded in a fully specified neighborhood in an observed network, the unit's program participation propensity can be evaluated based on a set of pre-selected network statistics. Having identified informative network statistics for a given problem, one can proceed with addressing program evaluation queries on networks.

4 Program Effect Evaluation: Focusing on the Participant Pool

The idea of using propensity score to remove bias in conventional, non-network based, treatment effect estimation has seen much success. However, in the presence of network effects, re-establishing the value of the idea requires care.

Under the non-inference assumption, the conventional average treatment effect (ATE) does not depend on the structure of the network, in which the studied units are embedded: this is because ATE is defined as the difference between the average outcomes in the situation where all the units are treated and the situation where none is (i.e. participant decision are pre-determined).Hence, For the same program, ATE should always turn out the same when the estimated for different networks. This observation provides good intuition for the reasons why causal inference theory on networks need to be revised: otherwise, since the treatment groups do not turn out different in the different networks (because of peer pressure), the respective ATE estimates would turn out different.

Continuing the decision of potential quantities of interest, observe that the network structure plays a significant role in defining the average treatment effect on the treated, ATT and on the untreated $ATUT$ (with $ATE = P(T = 1)ATT + P(T = 0)ATUT$, where $P(T = 1)$ is the probability of the treatment for a randomly selected unit in the population). As such, $ATT(ATUT)$ should turn out different for the same program on different network, because the program participant formation is in part driven by the units' connections. Note also, that since the participant pool depends on the network structure, ATT itself does not provide the same information about the program as in the case with independent participant decisions (where participant reason might be unknown but fixed for any individual). In other words, the success of a voluntary program in large part depends on the program's engagement/reach. Consider a program have a large average effect on the treated but engaging only a few potential participants. The same program disseminated on a different network, or even advertised differently on the same network may be end up having a smaller average effect on the treated while engaging a larger part of population. This motivates to incorporate network

covariate in propensity score. The estimands, we are going to focus on in this research are
(1) $ATE = E_u(Y_u^1 - Y_u^0)$ and
(2) $ATT = E_u(Y_u^1 - Y_u^0 | T_u = 1)$
To estimate ATE and ATT, we need to redefine the propensity score which we called 'expaned propenstiy score'. It is explained in the following section in detail.

5 Expanding the Reach of Propensity Score Concepts to Network Settings

The estimation of network-dependent program effects can be approached by expanding the concept of propensity score, i.e. explicitly incorporating network effects into it. The foundation of the expanded propensity score theory begins with its definition.

Definition 1 (Expanded propensity score) For a given u, the expanded propensity score is defined as the probability that u receives treatment, given its individual covariate vector X_u and set N_u that includes the treatment indicators and covariates of all the units in the local neighborhood of u,

$$e_u = e_u(X_u, N_u) = P(T_u = 1 | X_u, N_u). \tag{1}$$

The revised assumption of strong ignitability of treatment, originally started in the seminal work by Rosenbaum and Rubin (1983), can now be formulated.
Assumption 1 (Strong Ignorability): Consider a population of all the units under study embedded into a network with a fixed structure. For any feasible vector of observed neighborhood variable value N_u assume,

$$(Y_u^1, Y_u^0) \coprod T_u | (X_u, N_u), \tag{2}$$

$$\text{and } 0 < P(T_u = 1 | X_u, N_u) < 1. \tag{3}$$

Expression 2 implies that the outcomes of any unit u are independent of its treatment, given the unit's observed covariates values and its observed local neighborhood variable values. The symbol '$\coprod$' signifies conditional independence (Dawid, 1979). This means that the observed covariates of u together with the observed treatment indicators and covariate of all the u's neighbors (the units in the local neighborhood of u) include all the information, dependent on the treatment assignment T_u, that have causal effects on the response Y_u^1 and Y_u^0. Expression 3 implies that each unit with a given set of its covariate and neighborhood variable values is assumed to have a positive probability of appearing in either the treatment group or control group.
Theorem 1 (Unbiased Estimation of ATE from a large Sample) Under assumption 1, one has

$$\begin{aligned} E_u(Y_u^1 | e_u(X_u, N_u), T_u = 1) &= E_u(Y_u^1 | e_u(X_u, N_u)) \\ E_u(Y_u^1 | e_u(X_u, N_u), T_u = 0) &= E_u(Y_u^0 | e_u(X_u, N_u)) \end{aligned}$$

and hence,

$$\begin{aligned} E_{e_u(X_u,N_u)} \left(E_u(Y_u^1 | e_u(X_u, N_u), T_u = 1) - E_u(Y_u^0 | e_u(X_u, N_u), T_u = 0) \right) = \\ E_{e_u(X_u,N_u)} \left(E_u \left(Y_u^1 | e_u(X_u, N_u) \right) - E_u \left(Y_u^0 | e_u(X_u, N_u) \right) \right) = E_u \left(Y_u^1 | e_u(X_u, N_u) \right) \end{aligned} \tag{4}$$

giving the ATE. Similarly, by identifying the units from the control pool whose expanded propensity score values most closely match the expected propensity score values of each of the treated or untreated units, one will arrive at the unbiased estimators of ATT.

The expanded propensity score generalized the conventional propensity score of Rosenbaum and Rubin (1983) by incorporating the explicit unit dependency structures captured as N_u. Importantly, if no network-induced or other types of dependence exists between units, then (1) reduces to the definition of the conventional propensity score. But in general, the expanded propensity score of unit u is defined to be a function of N_u, which includes the treatment indicators and covariates in the local neighborhood of N_u. Thus, different units' propensity scores are dependent.

In order to show how the expanded propensity score can be estimated and used, a feasibility study is reported next. Its description begins with the overview of the data obtained over the course of the Framingham Heart Study.

6 Analysis of the Framingham Heart Study Data

6.1 FHS Overview

The Framingham Heart Study (FHS) data is a well-known dataset, which contains social network information as well as a large number of individual covariates characterizing the study participants. The data for the study have been collected since 1948. The study was initiated with 5,209 people in Framingham, MA, who were enrolled into the 'Original Cohort' (Dawber, 1980). In 1971, the 'Offspring Cohort', composed of many of the children of the Original Cohort, and their spouses, was enrolled (Feinleib et al., 1975). The examinations took place during 3-year periods centered in 1973, 1981, 1985, 1989, 1992, 1997, and 1999 (Christakis and Fowler, 2007). This cohort of 5,124 people has had almost no loss to follow-up other than due to death. Christakis and Fowler (2013) present multiple statistics of the offspring cohort. In 2002, enrollment of the so-called 'Third Generation Cohort' began, consisting of 4,095 of the children of the Offspring Cohort. The study has also involved certain other smaller cohorts. Participants in all the cohorts come to a central facility for detailed physical examinations and data collection every 24 years.

For many decades, the FHS has maintained handwritten tracking sheets that administrative personnel have used to identify people close to participants for the purpose of facilitating follow-up. These hand written documents were migrated into a digital database later on. Based on these database, the social network of participants was constructed.

The FHS recorded complete information about all the first-order relatives (parents, spouses, siblings, children), whether alive or dead, and also about at least one "close friend" of each participant (using the questions like "please tell us the name of a close friend, to whom you are not related", "with whom you are close enough that they would know where you are if we can't find you"). This information was collected at each of the seven exams between 1971 and 2003. Detailed home address information was also captured at each time point. The information about place of employment at each wave allowed for the identification of ties to coworkers within the network (by observing whether two people worked at the same place at the same time).

Over the course of the follow-up, the participants spread out across the USA, but they nevertheless continued to participate in the FHS. As a person's family changed because of birth, death, marriage, or divorce, and as their contacts changed because of residential moves, new places of employment, or new friendships, this information was captured. For any given 'ego' (the person of interest) in the data, a particular 'alter' (a person who has a relationship with the ego) may usually be placed in one category: spouse, sibling, parent, friend, and so on; although, sometimes, multiple categories

were allowed (for example, a coworker or neighbor might be a friend or sibling).

In the offspring cohort, any person to whom these subjects were linked in any sort of relationship in any of the FHS cohorts, including the Offspring Cohort itself, can serve as alters. In total, there were 12,067 egos and alters across all the cohorts of the FHS who were connected at some point during 1971 to 2003. There are ties among the individuals both inside and outside the sample. For example, an ego might be connected to two siblings, one of whom was also a participant in the FHS and one of whom was not. For those who were also participants, we could observe their attributes (for example, their health status) longitudinally. Overall, as of 2009 and wave 7 of data collection, there were 53,228 observed familial and social ties to the 5124 subjects observed at any time from 1971 to 2009, yielding an average of 10.4 ties per subject within the network (not including ties to residential neighbors). Fully 83% of subjects' spouses were also in the FHS and 87% of subjects with siblings had at least one sibling in the FHS.

Christakis and Fowler (2007) identified spouses, siblings, and other contacts who are outside our sample using the basic information about them; and although they are not in the FHS. Importantly, 45% of the 5124 subjects were connected via friendship to another person in the FHS at some point, which allowed them to observe outcomes for both the naming friend and the named friend. In total, there were 3542 such friendships for an average of 0.7 friendship ties per subject. For 39% of the subjects, at least one coworker was captured in the network at some point. For 10% of the subjects, an immediate (nonrelative) residential neighbor was also present (more expansive definitions, such as living within 100m, resulted in many more subjects having identifiable 'neighbors').

FHS implemented a variety of changes to the data to help protect subject confidentiality, however, before posting. Specifically: (1) all date information was changed to a monthly resolution rather than daily; (2) only 9000 cases rather than 12,000 could be posted (e.g., all non-genetically related relative ties such as adopted siblings, step-children, etc., were removed); (3) individuals who did not consent to the release of 'sensitive information' were excluded; and (4) the available covariates (e.g., geographic coordinates) were restricted. This dataset is distributed via the SHARE database at dbGAP (http://www.ncbi.nlm.nih.gov/gap).

6.2 Data Preprocessing

The present study relied on the "Original and Children cohort" data from FHS-Net. We found that the social network between the original cohort participants must not have been documented properly: the network is too sparse. The social network between the "Children Cohort" participants, however, appeared to be quite realistic. We included age and marital status as individual covariates (X) in our study. We calculated BMI using weight and height which is used as Response in our analysis. First, we checked for any missing covariate values, and removed the individuals with missing values in their records from the dataset. We ended up with 2,280 eligible individuals. While the FHS allows for building a directed network (because of the edge data were collected), we chose to work with an undirected network, assuming that the reported friendships must be mutual and the lack of reciprocity is attributed to the fact the study design did not encourage the identification of all social connections of a given individual. There resulting social network featured 7047 edges. We conducted descriptive analyses of the individual covariates; the results are found below.

1. Distribution of weight is positively skewed (Fig. 2a). But there are no outliers in the weight variable.

2. Height is approximately normally distributed (Fig. 2b).

3. In FHS, Smoker is a nominal variable with 3 levels. These 3 levels are: non-smoker, smoker and former smoker, labeled 0, 1 and 2, respectively (Fig. 2c). Note that, former smokers are the individuals who have not smoked for at least one year. In the considered population, there are 39.82% non-smokers, 40.31% smokers and 19.87% former smokers (Table 1). We decided to merge the smoker and former smoker levels into a single level for better interpretability. The smoker variable serves as a treatment variable (T) in our study.

4. Marital status is also a nominal variable with 5 levels. These levels are: single, married, widowed, divorced and separated, labeled 1, 2, 3, 4 and 5, respectively (Fig. 2d). We find that 79.96% of the individuals in the considered population are married (see Table 2) and the respective frequencies of the other levels are very low. We decided to merge those levels into a single level - unmarried.

5. The male and female counts in the population are about the same (Table 3) (Fig. 2e).

6. Age is approximately normally distributed (Fig. 2f).

7. Finally, we calculated the Body mass index (BMI) of all the individuals in the considered population using their heights and weights. The BMI variable serves as a response variable (Y) in our study.

Note that since the height and weight were used to compute the response (BMI), these variables were not used as covariates; thus, the covariates (X) are age, sex, and marital status.

The prepared FHS-Net dataset was too large for our inference methods to be directly applied (ERGM parameter learning would be both memory- and time-consuming); we then decided to apply a clustering algorithm to partition the network into cohesive subcommunities.

Table 1: Frequency table of smokers

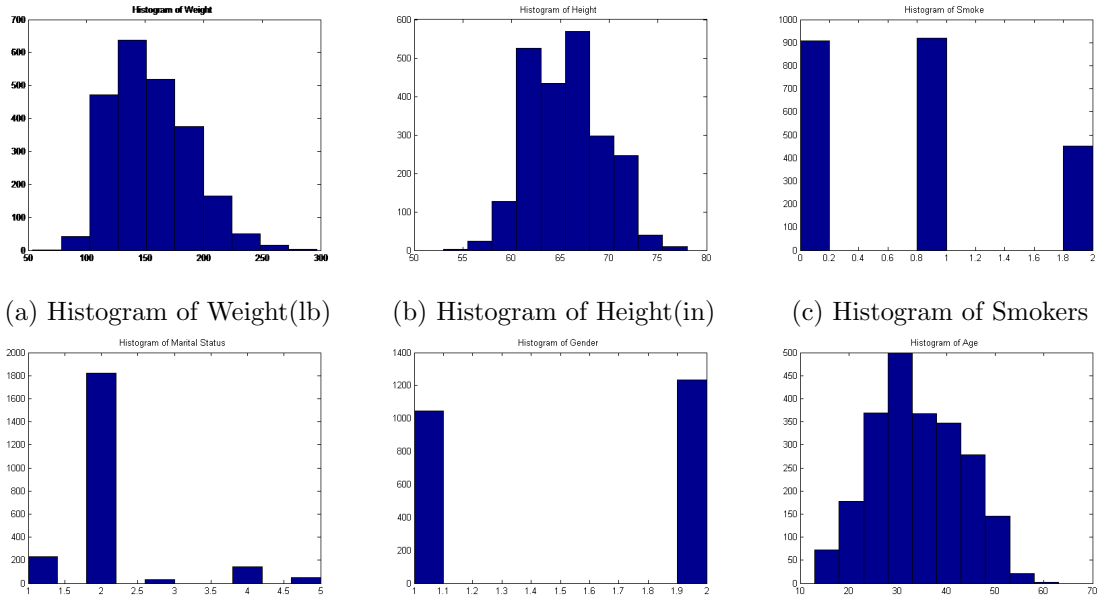
Level	Value	Count	Percent
Non-smoker	0	908	39.82%
Smoker	1	919	40.31%
Former smoker	2	453	19.87%

Table 2: Frequency table of Marital Status

Level	Value	Count	Percent
Single	1	231	10.13%
Married	2	1823	79.96%
Widowed	3	35	1.54%
Divorce	4	144	6.32%
Separated	5	47	2.06%

Table 3: Frequency table of Gender

Level	Value	Count	Percent
Male	1	1044	45.79%
Female	2	1236	54.21%



(a) Histogram of Weight(lb) (b) Histogram of Height(in) (c) Histogram of Smokers

(d) Histogram of Marital status (e) Histogram of Gender (f) Histogram of Age

Figure 2: Histogram of different attribute of FHS-Net

6.3 Network Clustering

Network clusters in social network analysis are defined as "cohesive" groups of vertices, i.e., with the edge density within the groups being higher than between them. The problem of detecting such communities within networks is well studied. Early approaches including the KernighanLin algorithm (Kernighan and Lin, 1970), spectral partitioning (Fiedler, 1973; Pothen et al., 1990) and hierarchical clustering (Scott, 1988) worked well for specific types of problems (particularly graph bisection or problems with well-defined vertex similarity measures), but were found to perform poorly in more general cases (Newman, 2004a).

A number of newer, network clustering algorithms have been proposed by Girvan and Newman (Girvan and Newman, 2002; Newman and Girvan, 2004). They first proposed a divisive algorithm that uses edge betweenness as a metric to identify the boundaries of communities. This algorithm has been applied successfully to a variety of networks, including networks of email messages, human and animal social networks, networks of collaborations between scientists and musicians, metabolic networks and gene networks (Newman and Girvan, 2004; Gastner and Newman, 2004; Guimera et al., 2003; Holme et al., 2003; Holme and Huss, 2003; Tyler et al., 2005; Boguná et al., 2003; Wilkinson and Huberman, 2004; Arenas et al., 2004). However, as noted in (Newman and Girvan, 2004), the algorithm is quite inefficient computationally, running in $O(m^2n)$ time on an arbitrary network with m edges and n vertices, or $O(n^3)$ time on a sparse graph (one in which $m \backsim n$, which covers most real-world networks of interest). This restricts the algorithm's use to networks of at most a few thousand vertices with current hardware. Later, Newman and Girvin proposed an algorithm based on the greedy optimization of the quantity known as modularity (Newman and Girvan, 2004). This method appears to work well both in contrived test cases and in real-world situations, and is substantially faster than the original algorithm in (Holme et al., 2003). A naive implementation runs in time $O((m+n)n)$, or $O(n^2)$ on a sparse graph. Clauset et al. (2004) propose a new algorithm that performs the same greedy optimization as the algorithm of Newman (2004b)

with tree based data structure (Clauset et al., 2004). It runs far more quickly, in time $O(md \log n)$ where d is the depth of the Dendogram describing the network's community structure. This is not merely a technical advance but has substantial practical implications for application on extremely large networks. We decided to use this algorithm - the "Modularity"-based clustering algorithm - for our study.

106 clusters were detected by the algorithm in the prepared network. We ignored the clusters containing less than 20 nodes (such small sub-networks could not produce any meaningful results). We also calculated number of edges within each cluster and the number of edges between each cluster and rest of the clusters. We found the proportion of edges within the clusters to lie in the range 51-72% (see Table 4), which is acceptable given that the number of nodes outside any cluster is significantly greater than that inside it.

Table 4: Summary of clusters of FHS-Net identified by Modularity based clustering

Cluster	Number of node	Number of Edge within cluster	Number of edge from one cluster to rest of the clusters	Percent of edge within cluster
1	294	746	511	59.35%
2	290	1024	621	62.25%
3	271	694	611	53.18%
4	243	760	553	57.88%
5	207	909	583	60.92%
6	103	203	185	52.32%
7	101	170	162	51.20%
8	66	98	93	51.31%
9	60	91	55	62.33%
10	59	79	57	58.09%
11	50	75	59	55.97%
12	47	57	47	54.81%
13	23	31	12	72.09%

Note that social networks used in research studies are typically partially observed. Online social network site data (e.g. Facebook, google+, twitter etc.) cannot account for individuals who do not have online accounts or are inactive. The social networks collected via questionnaire are restricted to survey responders. Recently, Onnela and Christakis (2012) studied the spreading paths in partially social network networks (Onnela and Christakis, 2012). Although partial observations were found to inflate the length of the shortest path in a network (its diameter), the lengths of diffusion spread paths (i.e., trajectories of processes on networks) were not quite as inflated. Importantly, Onnela and Christakis (2012) shoed that cluster properties depend on the way in which clusters are discovered; then, they experimentally found that modularity-based clustering algorithms produces most desirable clusters.

On the final note, observe that the clusters identified in the FHS-net network by the modularity algorithm are quite large, with the sizes of the largest two clusters at 296 and 290 (see Table 4): such large sub-networks can reasonably well represent the whole network.

7 Model Design and Expanded Propensity Score Estimation

We employ the model for the expanded propensity score that combines logit and ERGM functions. The logit function captures the dependence of the treatment selection on the individual covariates, while the ERGM function captures the dependence of the treatment selection on the network structure. Parameterizing ERGMs can be quite challenging; to this end, we use a Markov Chain Monte Carlo (MCMC) sampling procedure, which prescribes to construct a Markov chain that has the target distribution as its equilibrium distribution.

The proposed approach is similar to the treatment of other classes of ERGMs, for which the parameterisations have been accomplished successfully (Robins et al., 2007; Wang et al., 2006), and also, the treatment of actor-oriented models (Snijders, 2005, 2002; Snijders et al., 2010).

Denote the vector of the treatment indicators over all the individuals participating in a program by T, the vector of their covariate vectors by X, and the graph capturing the social network structure in the considered population by N; for any individual u, denote the treatment indicator value and the covariate vector by T_u and X_u, respectively. Also, let the network statistic value for statistic j and its corresponding ERGM parameter be denoted by f_j and θ_j, respectively. According to our proposed model, the probability of any given realization of the treatment indicators of all the individuals in the population, T, given their individuals' attributes and the network of their connections, can be expressed as

$$\begin{aligned} P(T|X,N) &= \frac{1}{Z} exp\left(\sum_{j\in J}\theta_j f_j(T,N)\right)\prod_{u=1}^{n} P(T_u|X_u) \\ &= \frac{1}{Z} exp\left(\sum_{j\in J}\theta_j f_j(T,N)\right)\prod_{u=1}^{n} P(T_u=1|X_u)^{T_u} P(T_u=0|X_u)^{1-T_u}, \end{aligned} \tag{5}$$

where J is the set of the network statistics of interest, and n is a total number of individuals in the population. We modeled the second term in equation (5) assuming that

$$logit(T_u) = w'X_u = \sum_i w_i X_{u,i}, \tag{6}$$

where w_i is the parameter corresponding to the covariate vector element X_i. Using the sigmoid function $\sigma(\cdot)$ commonly employed in logistic regression analyses, one has

$$P(T_u = 1|X_u) = \sigma(w'X_u). \tag{7}$$

Substituting equation (6) into equation (5), one obtains

$$P(T|X,N) = \frac{1}{Z} exp\left(\sum_{j\in J}\theta_j f_j(T,N)\right)\prod_{u=1}^{n}\left(\sigma(w'X_u)\right)^{T_u}\left(1-\sigma(w'X_u)\right)^{1-T_u}. \tag{8}$$

In this study, in order to facilitate the parameter learning, we learn the logit component parameters first (in separation from ERGM component) by maximizing the likelihood of the observed network using a steepest ascent algorithm. Then, we learn the parameters of the ERGM component, given the logit parameter values. We now express the probability that any given individual u is treated

(i.e., $T_u = 1$), given the treatment indicators in the rest of the population $T_u^c = T \setminus T_u$, all the covariate vector values X and the social network structure N:

$$P(T_u = 1|T_u^c, X, N) = \frac{P(T_u = 1, T_u^c|X, N)}{P(T_u^c)}. \tag{9}$$

Similarly, one can write

$$P(T_u = 0|T_u^c, X, N) = \frac{P(T_u = 0, T_u^c|X, N)}{P(T_u^c)}. \tag{10}$$

Dividing equation (9) by equation (10),

$$\begin{aligned}
\frac{P(T_u = 1|T_u^c, X, N)}{P(T_u = 0|T_u^c, X, N)} &= \frac{P(T_u = 1, T_u^c|X, N)}{P(T_u = 0, T_u^c|X, N)} \\
&= \frac{exp(\sum_{j\in J} \theta_j f_j(T_u = 1, T_u^c, N)) \cdot \sigma(w'X_u)}{exp(\sum_{j\in J} \theta_j f_j(T_u = 0, T_u^c, N)) \cdot (1 - \sigma(w'X_u))} \\
&= exp\left(\sum_{j\in J} \theta_j \left(f_j(T_u = 1, T_u^c, N) - f_j(T_u = 0, T_u^c, N)\right)\right) exp(w'X_u). \qquad (11)
\end{aligned}$$

Note that in (11), because both the numerator and denominator are expressed using the same model (8), the normalizing constant Z cancels out. Now, by the law of total probability,

$$P(T_u = 1|T_u^c, X, N) + P(T_u = 0|T_u^c, X, N) = 1. \tag{12}$$

Hence, the expanded propensity scores can be derived explicitly by solving equations (11) and (12).

Two points pertaining to (11) and (12) are worth emphasizing. First, note that the network statistics $f_j, j \in J$, depend on individual who are in the neighbor of u; thus, the expanded propensity score of unit u is computed with the understanding that a change in this unit's treatment would affect the treatment propensity of all the units in u's neighborhood (because u itself is in their neighborhoods). Second, if the dependencies between units are removed, then the only remaining ingredients in (11) are those related to u's covariates X_u; in this case, (12) reduces to the expression for the conventional propensity score.

We applied a parameter learning algorithm on the 13 largest clusters found by the modularity based clustering algorithm. First, we tried to estimate the logit parameters - the covariate-related ones - for all the 13 clusters (Table 5). We found that the logit parameters could not be estimated in clusters 9 and 11 (the likely reason for this is that the subpopulations within the clusters were not homogeneous in their smoking selection patterns); we thus excluded these clusters from further analysis. We applied an MCMC based parameter estimation algorithm on the remianing 11 clusters, with the model including the network statistics described in the three equations below. However, we observed degeneracy arising in learning on all the clusters. We then redefined our network statistics similar to the method proposed by Schweinberger (2011): he showed that degeneracy related to some unstable network statistics can be overcome by re-scaling unstable sufficient statistics with the help of minimum and maximum values of those sufficient statistics. Consequently, we re-scaled the network statistics by dividing them by the maximum possible values (F_1, F_2, F_3) of the respective

network statistics (f_1, f_2, f_3) (13, 14, 15).

$$f_1(T,N) = \frac{1}{F_1}\sum_{i=1}^{n} I(T_i = 1), \tag{13}$$

$$f_2(T,N) = \frac{1}{2F_2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(T_i = T_j)N_{ij}, \text{ where } (i \neq j), \tag{14}$$

$$f_3(T,N) = \frac{1}{3F_3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I(T_i = T_j)I(T_i = T_k)N_{ij}N_{jk}N_{ik}, \text{ where } (i \neq j \neq k), \tag{15}$$

where $F_1 = n$, $F_2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} N_{ij}$ and $F_3 = \frac{1}{3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} N_{ij}N_{ik}N_{jk}$.

With the redefined network statistics, the model parameters were satisfactorily estimated for all the 11 clusters (Table 5). Then we conducted a Goodness of Fit analysis of all the parameters, as proposed by Hunter et al. (2008), by simulating a predetermine number of networks and checking how well the learned models captured other network statistics (not used in modeling). For such a statistic, we used Treated K-Star counts; for node i, this count is

$$K(i,1) = \sum_{j \neq i} N_{ij}T_iT_j. \tag{16}$$

Of all the clusters (Fig. 3), the Treated K-star count could not reproduced only in cluster 13 (see Fig. 3k in the Appendix). We also analyzed the fitted means of network statistics with respect to the observed network statistics. We found that the mean network statistics converge very quickly - within first 5 iterations (Fig. 4, 5, 6), and then the means slightly fluctuate which is common in stochastic optimization. We also find out that the converged means of the network statistic are closer to the observed values in smaller clusters (Fig. 5, 6), which is a good sign. However, there is a gap between the observed network statistics and the converged means for larger clusters (Fig. 4): this is also is expected in parameter learning and in application of the Newton-Ralphson algorithm, with parameter estimate convergence becoming slow when current estimates are close to the truth. We concluded that more MCMC samples could be used to achieve a better fit for larger clusters. However, since this could be time consuming; besides, this problem can be tackled by finding a good initial point for the algorithm in the parameter space.

8 Estimation of the Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT)

Two popular conventional propensity score-based approaches for estimating ATE and ATT are 1) Propensity Stratification (PS), and 2) Inverse Propensity (IP) approaches. Both these methods have their advantages and deficiencies.

8.1 Propensity Score Stratification based approach

Stratification based on estimated propensity scores and comparison of treated and non-treated within the strata is a common analytic strategy. The justification for conducting stratification following greedy matching is provided by Rosenbaum and Rubin (1983). The process is intuitive:

Table 5: Parameter Estimation for different clusters of FHS-Net

Cluster	Number of nodes	Number of Edges	Logit Parameters				ERGM Parameters		
			w_0	w_1	w_2	w_3	θ_1	θ_2	θ_3
1	294	746	-0.118	0.547	-0.517	1.259	0.861	-0.254	-0.022
2	290	1024	-0.150	0.631	-0.510	0.437	7.737	-0.243	0.251
3	271	694	0.028	0.242	0.051	0.792	0.939	-0.222	-0.082
4	243	760	0.358	-0.181	-0.672	2.589	1.345	0.118	0.018
5	207	909	0.928	0.487	-0.977	-0.770	1.762	-0.047	0.667
6	103	203	-1.840	0.806	0.616	4.188	1.283	-0.512	-0.065
7	101	170	-0.454	0.249	-0.444	1.975	0.535	-0.244	-0.161
8	66	98	0.260	-0.415	-0.596	2.980	0.872	-0.318	0.146
9	60	91							
10	59	79	-0.828	1.417	-0.363	0.436	0.843	-0.481	-0.551
11	50	75							
12	47	57	10.106	-8.532	-0.551	-0.688	2.497	-0.434	1.780
13	23	31	-2.372	0.872	0.436	4.663	0.958	-0.572	-1.567

sort the sample into K strata based on the estimated propensity scores, calculate the mean difference of outcomes and variance of differences between treated and control participants within each stratum, estimate the mean difference (ATE) for the whole sample and test whether or not the sample difference is statistically significant. Assume that the propensity score range is divided into K strata, with the propensity score of the k-th stratum denoted by e_k, the number of the treated observations in the k-th stratum denoted by $n(e_k, T = 1)$, the number of control observations in the k-th stratum denoted by $n(e_k, T = 0)$, and the number of observations in the k-th stratum denoted by $n(e_k)$. Then, the probability that an observation belongs to the k-th stratum can be expressed as

$$P(e_k) = \frac{n(e_k)}{\sum_{i=1}^{K} n(e_i)} = \frac{n(e_k)}{N}, \tag{17}$$

where N is a total number of observations. Similarly, the probability that a treated observation belongs to the k-th stratum is

$$P(e_k, T = 1) = \frac{n(e_k, T = 1)}{\sum_{i=1}^{K} n(e_i, T = 1)} = \frac{n(e_k, T = 1)}{N_1}, \tag{18}$$

where N_1 is a total number of the treated observations. With the treated response and the untreated response denoted by Y^1 and Y^0, respectively, the ATE estimator is given,

$$\begin{aligned} Y &= Y^1 - Y^0 \\ \bar{Y} = E(Y) &= E[Y^1 - Y^0] \\ &= E_{e(x)}\left[E\left[Y^1|e(x)\right] - E\left[Y^0|e(x)\right]\right] \\ &= \sum_{k=1}^{K} P(e_k)\left[\bar{Y^1_{e_k}} - \bar{Y^0_{e_k}}\right] \\ &= \sum_{k=1}^{K} \frac{n(e_k)}{N}\left[\bar{Y}^1_{e_k} - \bar{Y}^0_{e_k}\right]. \end{aligned} \tag{19}$$

The variance of the ATE estimator is

$$\begin{aligned} var(\bar{Y}) &= \sum_{k=1}^{K} \left(\frac{n(e_k)}{N}\right)^2 var\left(\bar{Y}_{e_k}^1 - \bar{Y}_{e_k}^0\right) \\ &= \sum_{k=1}^{K} \left(\frac{n(e_k)}{N}\right)^2 \left(var(\bar{Y}_{e_k}^1) + var(\bar{Y}_{e_k}^0)\right) \\ &= \sum_{k=1}^{K} \left(\frac{n(e_k)}{N}\right)^2 \left(\frac{var(Y_{e_k}^1)}{n(e_k, T=1)} + \frac{var(Y_{e_k}^0)}{n(e_k, T=0)}\right). \end{aligned} \tag{20}$$

The estimator is asymptotically normally distributed. The significance test for the null hypothesis of "no treatment effect" can be performed based on the z statistic, with the critical value $z^* = \frac{\bar{Y}}{\sqrt{var(\bar{Y})}}$.

Similar to ATE, ATT estimator is obtained,

$$\begin{aligned} \bar{Y}_{T=1} &= E[(Y^1 - Y^0)|T=1] \\ &= E_{e(x)}\left[E\left[(Y^1 - Y^0)|T=1, e(x)\right]\right] \\ &= \sum_{k=1}^{K} P(e_k, T=1)\left[\bar{Y}_{e_k}^1 - \bar{Y}_{e_k}^0\right] \\ &= \sum_{k=1}^{K} \frac{n(e_k)}{N_1}\left[\bar{Y}_{e_k}^1 - \bar{Y}_{e_k}^0\right]. \end{aligned} \tag{21}$$

Its variance is

$$\begin{aligned} var(\bar{Y}_{T=1}) &= \sum_{k=1}^{K} \left(\frac{n(e_k, T=1)}{N_1}\right)^2 var\left(\bar{Y}_{e_k}^1 - \bar{Y}_{e_k}^0\right) \\ &= \sum_{k=1}^{K} \left(\frac{n(e_k, T=1)}{N_1}\right)^2 \left(var(\bar{Y}_{e_k}^1) + var(\bar{Y}_{e_k}^0)\right) \\ &= \sum_{k=1}^{K} \left(\frac{n(e_k, T=1)}{N_1}\right)^2 \left(\frac{var(Y_{e_k}^1)}{n(e_k, T=1)} + \frac{var(Y_{e_k}^0)}{n(e_k, T=0)}\right). \end{aligned} \tag{22}$$

The ATT estimator is also asymptotically normally distributed. The significance test for the null hypothesis of "No treatment effect on the treated" can be performed with the critical z statistic $z^* = \frac{\bar{Y}_{T=1}}{\sqrt{var(\bar{Y}_{T=1})}}$.

This method is easy to implement and very intuitive, but one of its key disadvantages is the assumption of overlapping support of the estimated propensity scores between the treated and untreated groups. When this assumption is violated, users obtain strata that contain either treated or untreated units but not both, and this happens most often in the strata for the lowest and/or highest propensity score values. This limitation can be overcome, as suggested by Crump et al. (2009), by using a formula to determine the exact proportion of observations to be trimmed, or using a rule of thumb for trimming - discarding all the units with the estimated propensity scores outside the range [0.1, 0.9].

8.2 Inverse Propensity score based approach

Propensity score may be used without matching to calculate ATE and ATT (Hirano and Imbens, 2001; Hirano et al., 2003; McCaffrey et al., 2004; Rosenbaum, 1987). Propensity score weighting aims to re-weigh treated and control participants to make them representative of the population of interest, e.g., as in Horvitz-Thompson estimators (Horvitz and Thompson, 1952). The crucial element of such analysis is the development of weights based on the estimated propensity scores. Propensity score weighting approach takes different amount of information from different observations. For estimating ATE, the weight of a given observation is defined as follows (Guo and Fraser, 2009),

$$w(T,x) = \frac{T}{\hat{e}(x)} + \frac{1-T}{1-\hat{e}(x)}. \tag{23}$$

Let I_1 and I_0 be the indices for treated and control observations, respectively. An estimator for ATE is then simply difference of the weighted averages of the treated and control responses,

$$\bar{Y} = \left[\sum_{j\in I_1} \frac{1}{\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_1} \frac{y_{1i}}{\hat{e}(x_i)} - \left[\sum_{j\in I_0} \frac{1}{1-\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_0} \frac{y_{0i}}{1-\hat{e}(x_i)} \ . \tag{24}$$

To simplify the presentation, define $l_1 = y_1/\hat{e}(x)$ and $l_0 = y_0/(1-\hat{e}(x))$. With these pieces of notation,

$$\bar{Y} = \left[\sum_{j\in I_1} \frac{1}{\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_1} l_{1i} - \left[\sum_{j\in I_0} \frac{1}{1-\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_0} l_{0i} \ . \tag{25}$$

The variance of this ATE estimator is

$$\begin{aligned} var(\bar{Y}) &= \left[\sum_{j\in I_1} \frac{1}{\hat{e}(x_j)}\right]^{-2} \sum_{i\in I_0} var(l_{1i}) - \left[\sum_{j\in I_0} \frac{1}{1-\hat{e}(x_j)}\right]^{-2} \sum_{i\in I_0} var(l_{0i}) \\ &= \left[\sum_{j\in I_1} \frac{1}{\hat{e}(x_j)}\right]^{-2} n_1 var(l_{1i}) - \left[\sum_{j\in I_0} \frac{1}{1-\hat{e}(x_j)}\right]^{-2} n_0 var(l_{0i}) \end{aligned} \tag{26}$$

The formulae for the estimation of ATT is similar to those for ATE: the only difference is in the applied weights. In estimating the ATT, the weight of an observation is defined as follows (Guo and Fraser, 2009),

$$w(T,x) = T + (1-T)\frac{\hat{e}(x)}{1-\hat{e}(x)}. \tag{27}$$

Then, the ATT estimate is found as the difference between the weighted averages of the treatment and control responses,

$$A\bar{T}T = \frac{1}{n_1}\sum_{i\in I_1} y_{1i} - \left[\sum_{j\in I_0} \frac{\hat{e}(x_j)}{1-\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_0} \frac{y_{0i}\cdot\hat{e}(x_i)}{1-\hat{e}(x_i)} \tag{28}$$

Defining $l = \frac{y_0\hat{e}(x)}{1-\hat{e}(x)}$ and substituting it into equation (28), one obtains

$$A\bar{T}T = \frac{1}{n_1}\sum_{i\in I_1} y_{1i} - \left[\sum_{j\in I_1} \frac{\hat{e}(x_j)}{1-\hat{e}(x_j)}\right]^{-1} \sum_{i\in I_0} l_i. \tag{29}$$

The variance of this ATT estimator is,

$$\begin{aligned} var(\bar{ATT}) &= \frac{1}{n_1^2}\sum_{i\in I_1} var(y_{1i}) - \left[\sum_{j\in I_1}\frac{\hat{e}(x_j)}{1-\hat{e}(x_j)}\right]^{-2}\sum_{i\in I_0} var(l_i) \\ &= \frac{var(y_0)}{n_1} - \left[\sum_{j\in I_1}\frac{\hat{e}(x_j)}{1-\hat{e}(x_j)}\right]^{-2} n_0 var(l_i). \end{aligned} \tag{30}$$

Again, the ATE and ATT estimators are found as the differences of the weighted averages of the responses over the treated and control observations. Both estimators are asymptotically normally distributed. The significance tests of the 'no treatment effect' null hypotheses can be performed based on this asymptotic property. The inverse propensity score is a creative idea and easy to implement. One of the main benefits of this method is that an overlapping support of the propensity score values for the treated and control units is no longer an issue. However, some recent studies suggested that this method may have its own limitations. Freedman and Berk (2008) conducted a number of simulation experiments focusing on ATE estimation, and found that propensity score weighting was only optimal under three circumstances,

1. when study participants are independent and identically distributed,
2. when selection is exogeneous, and
3. when selection equation is properly specified (i.e., with correct predictor variables and functional forms).

When these conditions are not observed, it was found that weighting is likely to increase random errors in the estimates. Indeed, it appears to bias the estimated causal parameter. Given these findings, Freedman and Berk (2008) suggested that investigators should exercise caution in implementing the weight procedure. They also warn that it rarely makes sense to use the same set of covariates in the outcome equation and selection equation that predicts propensity. Reflecting on both the development nature of the field and the uncertainty surrounding the validity of emerging procedures, Kang and Schafer (2007) showed that the use of inverse probabilities as weights is sensitive to mis-specification of the propensity score model when some estimate's propensities are small. Caution seems warranted in the use of propensity scores as sampling weights.

8.3 Imbalance Check in the Clusters (Model 1)

The main goal of estimating propensity score is to produce balance. It is often desirable to check covariate balance after matching on propensity. Haviland et al. (2007) developed the absolute standardized difference in covariate means. d_{xm} denotes the absolute standardized difference in covariate means. The value d_{xm} is estimated by the following formula

$$d_{xm} = \frac{|M_{xt} - M_{xc}|}{S_x} \text{ and } S_x = \sqrt{\frac{S_{xt}^2 + S_{xc}^2}{2}} \tag{31}$$

where M_{xt} and M_{xc} denote means of covariate X for treated and control group, respectively. S_{xt} and S_{xc} denote standard deviation of covariate X for treated and control group.

We stratified treated and control into 5 subclasses based on estimated propensity score. Then we performed covariate imbalanced check for each covariate in each strata. We used 0.50 as threshold

to identify imbalance for difference strata for different cluster. Imbalance in covariate is marked by bold font (Table 6, 7, 8, 9, 10, 11, 12). If there are no observation in either treated or control group in a strata, then we remove that strata.

We find out Marital Status in strata 1 and strata 2 are not balanced for cluster 1 (Table 6). Similarly, Age in strata 1 is not balanced. There are only one strata which has both control and treated observations for cluster 4 (Table 9, which is clear indication of imbalance. 7 out of 12 imbalance measures are more than 0.50 for cluster 5 (Table 9) and imbalance measure are also high. 9 out 15 imbalance measures are more than 0.50 for cluster 6. Similarly, we analyze imbalance for rest of the cluster (Table 11, 12). Other than cluster 4, 5 and 6 rest of the clusters have 2 to 4 imbalances measure more than 0.50. We decided to remove cluster 4, 5 and 6 from further analysis. Rest of clusters are considered for ATE and ATT estimation discussed in the following section.

Table 6: Covariate Balance Check of Cluster 1 (Model 1)

Strata	Marital Status	Gender	Age
1	**0.80**	0.40	0.31
2	0.47	0.02	0.15
3	**0.52**	0.42	0.26
4	0.13	0.28	0.06

Table 7: Covariate Balance Check of Cluster 3 (Model 1)

Strata	Marital Status	Gender	Age
1	0.40	0.40	**0.54**
2	0.16	0.32	0.04
3	0.14	0.12	0.01
4	0.38	0.17	0.17

Table 8: Covariate Balance Check of Cluster 4 (Model 1)

Strata	Marital Status	Gender	Age
1	0.00	0.32	0.20

Table 9: Covariate Balance Check of Cluster 5 (Model 1)

Strata	Marital Status	Gender	Age
1	**1.00**	**1.24**	**3.00**
2	**0.63**	**2.30**	**2.45**
3	0.30	**0.58**	0.43
4	0.22	0.49	0.07

Table 10: Covariate Balance Check of Cluster 6 (Model 1)

Strata	Marital Status	Gender	Age
1	**1.00**	**1.00**	**1.23**
2	**0.82**	**0.82**	**3.27**
3	**1.13**	0.24	**1.13**
4	0.25	0.06	0.44
5	0.18	0.46	**0.73**

Table 11: Covariate Balance Check of Cluster 7 (Model 1)

Strata	Marital Status	Gender	Age
1	**0.63**	0.47	0.35
2	**0.88**	0.24	0.44
3	0.16	0.36	0.05
4	**1.71**	0.42	0.15

Table 12: Covariate Balance Check of Cluster 8 (Model 1)

Strata	Marital Status	Gender	Age
1	0.16	0.18	0.50
2	0.46	**0.89**	0.11

8.4 ATE and ATT Estimation in the Clusters (Model 1)

First, we calculated the estimates of ATE and ATT in all the eligible clusters using PS approach using 10 strata (Table 13). We find that the ATE and ATT estimates are positive in all 4 clusters. The significance test of ATE and ATT were conducted next; we find that the estimates are significantly different from zero in clusters 1 ($\alpha \leq 0.10$) (see Table 13 where the significant p-values and z-scores are marked with an asterisk).

We conducted a similar analysis employing the 'Inverse Propensity' based approach (see Table 14). Again, the ATE and ATT estimates are positive for all 4 clusters (Table 14). However, None of estimates are found to be significant due to high variance of estimates.

Note that in cluster 1 - the largest cluster with 296 nodes - the treatment effects are found to be positive and significant. ATE and ATT estimate for cluster 1 are 1.078 and 1.108 BMI units, respectively.

8.5 Non Parametric test for ATE (Model 1)

In the previous section we have performed parametric test to check significance of 2 different estimate of ATE. However, those parametric test are based on assumption that they are asymptotically normally distributed. If this assumption does not hold, then parametric test will have high Type 2 error. As a result these tests will be less powerful. It is very difficult to check asymptotic property.

Table 13: ATE and ATT based on Propensity Stratification approach

Cluster	ATE	ATT	Z-score for ATE	Z-score for ATT	P-value (ATE)	P-value (ATT)
1	**1.078**	**1.108**	2.167*	2.277*	0.015*	0.011*
3	0.308	0.263	0.564	0.451	0.286	0.326
7	0.663	0.683	0.869	0.869	0.192	0.192
8	0.540	0.559	0.673	0.688	0.251	0.246

Table 14: ATE and ATT based on Inverse Propensity approach

Cluster	ATE	ATT	Z-score for ATE	Z-score for ATT	P-value (ATE)	P-value (ATT)
1	0.910	1.046	0.575	0.547	0.283	0.292
3	0.331	0.202	0.173	0.094	0.431	0.463
7	0.815	0.810	0.402	0.315	0.344	0.376
8	0.248	0.468	0.116	0.184	0.454	0.427

Finally, we decided to conduct non-Parametric which it is more generalized. The significance test of the ATE may be performed by the Hodges-Lehmann aligned rank test (Hodges et al., 1962). Lehmann and D'Abrera (2006) described the test in detail (see Algorithm 1).

Result of Non parametric test for eligible clusters are summarized in Table 15. We find out that ATE for cluster 1 is significant. We find out similar result for Propensity stratification based approach (Table 13). P-value is less than 0.01 for non parametric test.

Table 15: Non parametric test of ATE (Model 1)

Cluster	ATE	Z score for ATE	P-value (ATE)
1	**1.078**	2.647*	0.004*
3	0.308	0.846	0.199
7	0.663	1.022	0.153
8	0.540	0.218	0.414

8.6 Modified Model for Expanded Propensity Score (Model 2)

As we have performed clustering over the whole social network and used each cluster for our analysis, there are "lost" edges excluded from our analysis. We then modified our network statistics so as to directly account for the "lost" edges. Let the vector of the decisions (treatment indicators) of the individuals in cluster C̈ be denoted by T, the decisions of the individuals who are not in cluster C̈ but connected to the individuals in C̈ be denoted by (T^0), the covariates of the individuals in C̈ be denoted by X, and the social network by denoted by N. Similar to equation (8), the conditional

Algorithm 1 : Hodges-Lehmann Algorithm

1. Compute the mean of the outcome for each matched stratum i and then create a centering score for each participant by subtracting the stratum's mean from the observed value of the outcome.
2. Sort the whole sample by the centering scores in an ascending order and then rank the scores; the ranked score is called aligned rank and is denoted as $k_{ij}(j = 1, ..., N_i)$, where i indicates the ith stratum, j indicates jth observation within the ith stratum, and N_i the total number of participants in the ith stratum.
3. For each stratum i, compute

$$k_{i.} = \frac{k_{i1} + \cdots + k_{iN_i}}{N_i} \text{ and } E\left(\hat{W}_s^{(i)}\right) = n_i k_{i.},$$

$$\text{and } Var\left(\hat{W}_s^{(i)}\right) = \frac{m_i n_i}{N_i(N_i - 1)} \sum_{j=1}^{N_i} (k_{ij} - k_{i.})^2$$

   where n_i the number of treated participant in ith stratum, m_i the number of control participant in ith stratum.
4. Across strata, calculate

$$\hat{W}_s = \sum_i \hat{W}_s^{(i)},\ E\left(\hat{W}_s\right) = \sum_i E\left(\hat{W}_s^{(i)}\right),\ \text{and } Var\left(\hat{W}_s\right) = \sum_i Var\left(\hat{W}_s^{(i)}\right),$$

   where $\hat{W}_s^{(i)}$ is the sum of the aligned ranks for the treated participants within the ith stratum. Note that the subscript s in the all the preceding equations indicates treatment participants.
5. Finally, calculate the following test statistics $z*$:

$$z^* = \left[\hat{W}_s - E\left(\hat{W}_s\right)\right] / \sqrt{Var\left(\hat{W}_s\right)}.$$

   The z^* statistics follows a standard normal distribution. Using z^*, the analyst can perform a significance test of a non directional hypothesis (i.e., perform a two tailed test) or a directional hypothesis (i.e. perform a one-tailed test).

distribution of T given X, N, T^0 can be expressed as

$$\begin{aligned} P(T|X,N,T^0) &= \frac{1}{Z} exp\left(\sum_{j\in J}\theta_j f_j\left(T,N,T^0\right)\right)\prod_{u=1}^{n} P(T_u|X_u) \\ &= \frac{1}{Z} exp\left(\sum_{j\in J}\theta_j f_j\left(T,N,T^0\right)\right)\prod_{u=1}^{n} P(T_u=1|X_u)^{T_u} P(T_u=0|X_u)^{1-T_u}. \quad (32) \end{aligned}$$

Now, using the index C for any node within cluster C̈ and the index C' for any node not in C̈ but connected to a node in C̈, define a new set of network statistics as follows,

$$f_1(T,N,T^0) = \frac{1}{F_1}\sum_{i\in C} I(T_i=1), \quad (33)$$

$$f_2(T,N,T^0) = \frac{1}{F_2}\left(\frac{1}{2}\sum_{i\in C}\sum_{j\in C'} I(T_i=T_j)N_{ij} + \sum_{i\in C}\sum_{j\in C'} I(T_i=T_j)N_{ij}\right). \quad (34)$$

$$\begin{aligned} f_3(T,N,T^0) &= \frac{1}{3F_3}\sum_{i\in C}\sum_{j\in C}\sum_{k\in C} I(T_i=T_j)I(T_i=T_k)N_{ij}N_{jk}N_{ik} \\ &+ \frac{1}{2F_3}\sum_{i\in C}\sum_{j\in C'}\sum_{k\in C'} I(T_i=T_j)I(T_i=T_k)N_{ij}N_{jk}N_{ik} \quad (35) \\ &+ \frac{1}{2F_3}\sum_{i\in C}\sum_{j\in C}\sum_{k\in C'} I(T_i=T_j)I(T_i=T_k)N_{ij}N_{jk}N_{ik}. \quad (36) \end{aligned}$$

where $F_1 = \sum_{i\in C} I(T_i=1)$, $F_2 = \frac{1}{2}\sum_{i\in C}\sum_{j\in C} N_{ij} + \sum_{i\in C}\sum_{j\in C'} I(T_i=T_j)N_{ij}$ and $F_3 = \frac{1}{3}\sum_{i\in C}\sum_{j\in C}\sum_{k\in C} N_{ij}N_{jk}N_{ik} + \frac{1}{2}\sum_{i\in C}\sum_{j\in C'}\sum_{k\in C'} N_{ij}N_{jk}N_{ik} + \frac{1}{2}\sum_{i\in C}\sum_{j\in C}\sum_{k\in C'} N_{ij}N_{jk}N_{ik}$. Here, f_1 represents the ratio of the number of individuals in C̈ that are treated to the number of individuals in C̈; f_2 represents the ratio of the number of 1-hop friend agreements both in C̈ outside of C̈ to the total number of edges within C̈ plus the number the edges connected to C̈; f_3 represents the ratio of the number of agreement-cliques (of size 3) among the individuals within C̈ and the individuals connected to C̈ to the number of cliques of size 3 among the individuals in C̈ and those connected to C̈.

We learned the parameters for this new model (32), (33), (34) and (36), and checked the goodness of fit statistics (Fig. 11) and convergence in mean of network statistics (Fig. 12, 13, 14). We find that the mean of network statistics converge for all cluster. The goodness of fit results with the redefined network statistics is good for all the cluster except for cluster 7 (Fig. 11g), which was consequently excluded from further analysis.

8.7 Imbalance check for Model 2

We conducted imbalanced check for Model 2 similar to Model 1. We find out Marital Status in strata 1 and Age in Strata 2 are not balanced for cluster 1 (Table 16). Similarly, Marital status in strata 1, Gender in strata 3 and Age in strata 3 are not balanced. 3 out of 4 strata are not balanced for all 3 covariates for cluster 4 (Table 18) and imbalance score are high. So we decided

not to analyze this cluster in the future. There are only one strata which has both control and treated observations for cluster 5 (Table 19), which is a clear indication of imbalance. Similarly we analysis imbalanced rest of the cluster. Other than cluster 4 and 5, rest of the clusters has 2 to 4 imbalances. We decided to remove cluster 4 and 5 from further analysis. We calculate ATE and ATT for eligible clusters in the following section.

Table 16: Covariate Balance Check of Cluster 1 (Model 3)

Strata	Marital Status	Gender	Age
1	**0.94**	0.31	0.23
2	0.05	0.44	**0.67**
3	0.26	0.18	0.10
4	0.31	0.28	0.18
5	0.28	0.49	0.18

Table 17: Covariate Balance Check of Cluster 3 (Model 3)

Strata	Marital Status	Gender	Age
1	**0.53**	0.24	0.25
2	0.30	0.48	0.28
3	0.42	**0.93**	**0.65**
4	0.15	0.29	0.06
5	0.12	0.08	0.04

Table 18: Covariate Balance Check of Cluster 4 (Model 3)

Strata	Marital Status	Gender	Age
1	**0.82**	**1.00**	**1.81**
2	**1.22**	**0.71**	**1.27**
3	**0.58**	**1.72**	**1.73**
4	0.05	0.26	0.21

Table 19: Covariate Balance Check of Cluster 5 (Model 3)

Strata	Marital Status	Gender	Age
1	0.26	**0.52**	0.02

8.8 ATE and ATT Estimation in the Clusters (Model 2)

We calculated ATE and ATT using both Propensity Stratification (Table 23) and Inverse Propensity based approach (Table 24). ATE and ATT are positive for all cluster except cluster 6. ATE estimates are statistically significant for cluster 1 and none of the ATT is significant.

Table 20: Covariate Balance Check of Cluster 6 (Model 3)

Strata	Marital Status	Gender	Age
1	0.39	0.43	**1.00**
2	**1.32**	0.50	**0.67**
3	0.07	0.26	0.48
4	0.33	0.30	**0.73**

Table 21: Covariate Balance Check of Cluster 7 (Model 3)

Strata	Marital Status	Gender	Age
1	**0.63**	**0.67**	0.27
2	**0.51**	0.13	0.36
3	0.29	0.42	0.20
4	**1.14**	0.29	0.12

Table 22: Covariate Balance Check of Cluster 8 (Model 3)

Strata	Marital Status	Gender	Age
1	0.00	**1.00**	**1.41**
2	0.00	0.33	0.36
3	0.06	0.17	0.47
4	0.39	**1.00**	0.16

The ATE and ATT estimates are not significant for any cluster using IP (Table 24). We conclude that ATE and ATT of smoking on BMI are on average 0.6006 and 0.537. It means that if all the considered individuals in the 'Children Cohort' smoked, then their BMI would be 0.6006 units greater compared to the scenario where none of them smoked. Also, if the smokers in the 'Children Cohort' were not to smoke, their BMI would drop on average by 0.537 units.

Table 23: ATE and ATT calculation based on Propensity stratification for Model 3

Cluster	ATE	ATT	Z-score for ATE	Z-score for ATT	P-value (ATE)	P-value (ATT)
1	**0.850**	0.729	1.409*	1.091	0.079*	0.138
3	0.375	0.400	0.676	0.673	0.250	0.251
6	-1.225	-1.317	-1.146	-1.156	0.126	0.124
8	0.577	0.483	0.687	0.555	0.246	0.289

8.9 Non Parametric test for ATE (Model 2)

We conducted non parametric test for ATE which is summarized in Table 25. We find out very similar result to parametric test for the Model 2 (Table 23). We find out ATE is significant for

Table 24: ATE and ATT calculation based on Inverse propensity approach for Model 3

Cluster	ATE	ATT	Z-score for ATE	Z-score for ATT	P-value (ATE)	P-value (ATT)
1	1.033	1.161	0.282	0.291	0.389	0.385
3	1.373	0.884	0.173	0.200	0.431	0.421
6	-1.052	-0.461	-0.110	-0.101	0.456	0.460
8	0.109	0.461	0.042	0.153	0.483	0.439

cluster 1 (Table 25) using Hodges-Lehmann test. Conclusion of significance test are same for both non-parametric and parametric test (Table 23). But P-value of non-Parametric tests are different from parametric test.

Table 25: Non parametric test of ATE for Model 2

Cluster	ATE	Z score for ATE	P-value (ATE)
1	**0.850**	2.696*	0.004*
3	0.375	0.854	0.197
6	-1.225	-1.259	0.104
8	0.577	0.386	0.350

9 ATE and ATT Estimation without Blocking Network Effects

To conclude our study, we used the conventional propensity score-based methodology to estimate ATE and ATT without taking into account social network. In this case, propensity score is a function of individual covariates only. We calculate ATE and ATT for cluster 1, 3, 6, 7 and 8 to compare with the finding of the models including social network. We find out the ATE and ATT estimates are positive for all the clusters except cluster 6 (Table 26, 27). We find out the ATE and ATT are higher for Model 1 (Table 13) to estimate without considering social network (Table 26). Similarly, we find out that ATE and ATT are higher for Model 2 (Table 23) than estimate without considering social network (Table: 26).

Similar results are obtained using the IP based approach. ATE and ATT are positive for all clusters except cluster 6 (Table 27), and none of them is significant. It can be concluded that by ignoring the network effects, the causal inference quality noticeably degrades (compare Table 27 with Table 23).

10 Conclusion

The findings of this study are summarized as follows,

1. The study confirmed that stable-network ERGMs can be successfully used to express the dependencies between the problem variables with networks of 100-300 nodes.

Table 26: ATE and ATT calculation based on Propensity stratification (without social network)

Cluster	ATE	ATT	Z score (ATE)	Z score (ATT)	P-value (ATE)	P-value (ATT)
1	0.515	0.500	1.054	1.058	0.146	0.145
3	0.181	0.158	0.344	0.298	0.366	0.383
6	**-1.459**	**-2.058**	-1.851^{*}	-2.490^{*}	0.032^{*}	0.006^{*}
7	0.509	0.526	0.680	0.702	0.248	0.241
8	0.105	0.237	0.144	-0.332	0.443	0.370

Table 27: ATE and ATT calculation based on Inverse propensity approach (without social network)

Cluster	ATE	ATT	Z score for ATE	Z score for ATT	P-value (ATE)	P-value (ATT)
1	0.589	0.655	0.729	0.615	0.233	0.269
3	0.135	0.081	0.211	0.111	0.416	0.456
6	-1.318	-1.652	-0.615	-0.569	0.269	0.285
7	0.586	0.487	0.507	0.317	0.306	0.376
8	0.067	0.125	0.037	0.050	0.485	0.480

2. The choice of the causal estimator is very important for causal analyses with networks (e.g., the variance of ATE estimator based on the Inverse Propensity method turned out to be much higher than that based on the Propensity stratification method).

3. Non parametric and parametric significance tests have led to the same conclusion in the analysis of FHS data.

4. We calculated propensity based on two different models: the first model used the data only within each cluster, while the second model also used the data of the external, directly connected individuals. The propensity score values and causal estimates turned out to be very different between the two models suggesting that the influence of the external individuals is a significant factor.

5. Importantly, it was discovered that the estimates computed with blocking the network effects turned out different from those computed without blocking the network effects (see Tables 13 vs. 26). Moreover, the difference was about the same in all the clusters! This suggests that failure to block the network effects can compromise the accuracy of causal effect estimation in connected communities.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association 97*(457), 284–292.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 113–132.

Abadie, A. and G. Imbens (2002). Simple and bias-corrected matching estimators for average treatment effects.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Albert, R. and A.-L. Barabási (2002). Statistical mechanics of complex networks. *Reviews of modern physics 74*(1), 47.

Aral, S., L. Muchnik, and A. Sundararajan (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences 106*(51), 21544–21549.

Aral, S. and D. Walker (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science 57*(9), 1623–1639.

Arenas, A., L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimera (2004). Community analysis in social networks. *The European Physical Journal B-Condensed Matter and Complex Systems 38*(2), 373–380.

Bauman, K. E. and S. T. Ennett (1996). On the importance of peer influence for adolescent drug use: commonly neglected considerations. *Addiction 91*(2), 185–198.

Boguná, M., R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas (2003). Emergence of clustering, correlations, and communities in a social network model. *arXiv preprint cond-mat/0309263*.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology 2*(1), 113–120.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology 38*(1), 155–200.

Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys 22*(1), 31–72.

Carrell, S. E., R. L. Fullerton, and J. E. West (2008). Does your cohort matter? measuring peer effects in college achievement. Technical report, National Bureau of Economic Research.

Cho, W. K. T., J. J. Sauppe, A. G. Nikolaev, S. H. Jacobson, and E. C. Sewell (2011). An optimization approach to matching and causal inference. Technical report, Technical report, University of Illinois at Urbana–Champaign, Urbana, IL.

Christakis, N. A. and J. H. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine 357*(4), 370–379.

Christakis, N. A. and J. H. Fowler (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine 358*(21), 2249–2258.

Christakis, N. A. and J. H. Fowler (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine 32*(4), 556–577.

Clauset, A., M. E. Newman, and C. Moore (2004). Finding community structure in very large networks. *Physical review E 70*(6), 066111.

Contractor, N. S., S. Wasserman, and K. Faust (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review 31*(3), 681–703.

Cox, D. R. (1958). Planning of experiments.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, asn055.

da Veiga, P. V. and R. P. Wilder (2008). Maternal smoking during pregnancy and birthweight: a propensity score matching approach. *Maternal and child health journal 12*(2), 194–203.

David, O. and R. A. Kempton (1996). Designs for interference. *Biometrics*, 597–606.

Dawber, T. R. (1980). *The Framingham Study: the epidemiology of atherosclerotic disease*, Volume 84. Harvard University Press Cambridge.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–31.

Diamond, A. and J. S. Sekhon (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics 95*(3), 932–945.

Feinleib, M., W. B. Kannel, R. J. Garrison, P. M. McNamara, and W. P. Castelli (1975). The framingham offspring study. design and preliminary data. *Preventive medicine 4*(4), 518–525.

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal 23*(2), 298–305.

Fletcher, A. C., N. E. Darling, L. Steinberg, and S. Dornbusch (1995). The company they keep: relation of adolescents' adjustment and behavior to their friends' perceptions of authoritative parenting in the social network. *Developmental Psychology 31*(2), 300.

Fowler, J. H. and N. A. Christakis (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences 107*(12), 5334–5338.

Freedman, D. A. and R. A. Berk (2008). Weighting regressions by propensity scores. *Evaluation Review 32*(4), 392–409.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks 1*(3), 215–239.

Gastner, M. T. and M. E. Newman (2004). Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America 101*(20), 7499–7504.

Girvan, M. and M. E. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences 99*(12), 7821–7826.

Guimera, R., L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas (2003). Self-similar community structure in a network of human interactions. *Physical review E 68*(6), 065103.

Guo, S. and M. W. Fraser (2009). *Propensity score analysis: Statistical methods and applications*, Volume 12. Sage Publications.

Halloran, M. E. and C. J. Struchiner (1995). Causal inference in infectious diseases. *Epidemiology*, 142–151.

Handcock, M. S., D. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris (2003). statnet: An r package for the statistical modeling of social networks. *Web page http://www. csde. washington. edu/statnet*.

Haviland, A., D. S. Nagin, and P. R. Rosenbaum (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological methods 12*(3), 247.

Heckman, J. J. (2000). Microdata, heterogeneity and the evaluation of public policy. *Bank of Sweden Nobel Memorial Lecture in Economic Sciences, Stockholm, Sweden*.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. *Handbook of labor economics 3*, 1865–2097.

Heckman, J. J., L. Lochner, and C. Taber (1998). General equilibrium treatment effects: A study of tuition policy. Technical report, National Bureau of Economic Research.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review*, 107–111.

Herron, M. C. and J. Wand (2007). Assessing partisan bias in voting technology: The case of the 2004 new hampshire recount. *Electoral Studies 26*(2), 247–261.

Hinde, R. A. (1976). Interactions, relationships and social structure. *Man*, 1–17.

Hirano, K. and G. W. Imbens (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology 2*(3-4), 259–278.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hodges, J., E. L. Lehmann, et al. (1962). Rank methods for combination of independent experiments in analysis of variance. *The Annals of Mathematical Statistics 33*(2), 482–497.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association 81*(396), 945–960.

Holme, P. and M. Huss (2003). Discovery and analysis of biochemical subnetwork hierarchies. *arXiv preprint q-bio/0309011*.

Holme, P., M. Huss, and H. Jeong (2003). Subnetwork hierarchies of biochemical pathways. *Bioinformatics 19*(4), 532–538.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482).

Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association 103*(481).

Imai, K. (2005). Do get-out-the-vote calls reduce turnout? the importance of statistical methods for field experiments. *American Political Science Review 99*(02), 283–300.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika 87*(3), 706–710.

Jaccard, J., H. Blanton, and T. Dodge (2005). Peer influences on risk behavior: an analysis of the effects of a close friend. *Developmental psychology 41*(1), 135.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jacobs, B., J. Hartog, and W. Vijverberg (2009). Self-selection bias in estimated wage premiums for earnings risk. *Empirical Economics 37*(2), 271–286.

Jencks, C. and S. E. Mayer (1990). The social consequences of growing up in a poor neighborhood. *Inner-city poverty in the United States 111*, 186.

Kang, J. D. and J. L. Schafer (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 523–539.

Kernighan, B. W. and S. Lin (1970). An efficient heuristic procedure for partitioning graphs. *Bell system technical journal 49*(2), 291–307.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 604–620.

Lehmann, E. L. and H. J. D'Abrera (2006). *Nonparametrics: statistical methods based on ranks*. Springer New York.

Leskovec, J., D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani (2010). Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research 11*, 985–1042.

Lindquist, M. A. and M. E. Sobel (2011). Graphical models, potential outcomes and causal inference: Comment on ramsey, spirtes and glymour. *NeuroImage 57*(2), 334–336.

Luce, R. D. and A. D. Perry (1949). A method of matrix analysis of group structure. *Psychometrika 14*(2), 95–116.

McCaffrey, D. F., G. Ridgeway, and A. R. Morral (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods 9*(4), 403.

Monge, P. R. and N. S. Contractor (2003). *Theories of communication networks*. Oxford University Press.

Mouw, T. (2006). Estimating the causal effect of social capital: A review of recent research. *Annual review of sociology*, 79–102.

Newman, M. E. (2004a). Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems 38*(2), 321–330.

Newman, M. E. (2004b). Fast algorithm for detecting community structure in networks. *Physical review E 69*(6), 066133.

Newman, M. E. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical review E 69*(2), 026113.

Nikolaev, A. G., S. H. Jacobson, W. K. T. Cho, J. J. Sauppe, and E. C. Sewell (2013). Balance optimization subset selection (boss): an alternative approach for causal inference with observational data. *Operations Research 61*(2), 398–412.

Oakes, J. M. (2004). The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine 58*(10), 1929–1952.

Onnela, J.-P. and N. A. Christakis (2012). Spreading paths in partially observed social networks. *Physical Review E 85*(3), 036106.

Pothen, A., H. D. Simon, and K.-P. Liou (1990). Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications 11*(3), 430–452.

Reinisch, J. M., S. A. Sanders, E. L. Mortensen, and D. B. Rubin (1995). In utero exposure to phenobarbital and intelligence deficits in adult men. *Jama 274*(19), 1518–1525.

Robins, G., P. Pattison, Y. Kalish, and D. Lusher (2007). An introduction to exponential random graph (¡ i¿ p¡/i¿¡ sup¿*¡/sup¿) models for social networks. *Social networks 29*(2), 173–191.

Robins, G., T. Snijders, P. Wang, M. Handcock, and P. Pattison (2007). Recent developments in exponential random graph (¡ i¿ p¡/i¿*) models for social networks. *Social networks 29*(2), 192–215.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association 82*(398), 387–394.

Rosenbaum, P. R. (1989). Optimal matching for observational studies. *Journal of the American Statistical Association 84*(408), 1024–1032.

Rosenbaum, P. R. (1991). A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society. Series B (Methodological)*, 597–610.

Rosenbaum, P. R., R. N. Ross, and J. H. Silber (2007). Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *Journal of the American Statistical Association 102*(477), 75–83.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rubin, D. (2006). Matched sampling for casual effects. *AMC 10*, 12.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral statistics 2*(1), 1–26.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 34–58.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association 74*(366a), 318–328.

Rubin, D. B. (1980). Bias reduction using mahalanobis-metric matching. *Biometrics*, 293–298.

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 1213–1234.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association 100*(469).

Sampson, S. F. (1968). *A novitiate in a period of change: An experimental and case study of social relationships*. Ph. D. thesis, Cornell University, September.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association 106*(496), 1361–1370.

Scott, J. (1988). Social network analysis. *Sociology 22*(1), 109–127.

Scott, J. and P. J. Carrington (2011). *The SAGE handbook of social network analysis*. SAGE publications.

Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for r. *Journal of Statistical Software, Forthcoming*.

Snijders, T. A. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure 3*(2), 1–40.

Snijders, T. A. (2005). Models for longitudinal network data. *Models and methods in social network analysis 1*, 215–247.

Snijders, T. A., G. G. Van de Bunt, and C. E. Steglich (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks 32*(1), 44–60.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association 101*(476), 1398–1407.

Tu, W. and X.-H. Zhou (2002). A bootstrap confidence interval procedure for the treatment effect using propensity score subclassification. *Health Services and Outcomes Research Methodology 3*(2), 135–147.

Tyler, J. R., D. M. Wilkinson, and B. A. Huberman (2005). E-mail as spectroscopy: Automated discovery of community structure within organizations. *The Information Society 21*(2), 143–153.

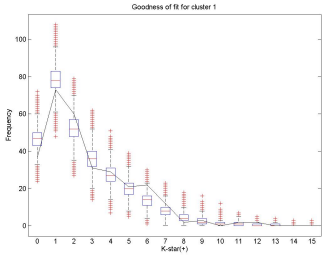
Wang, P., G. Robins, and P. Pattison (2006). Pnet: a program for the simulation and estimation of exponential random graph models. *University of Melbourne*.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of small-worldnetworks. *nature 393*(6684), 440–442.

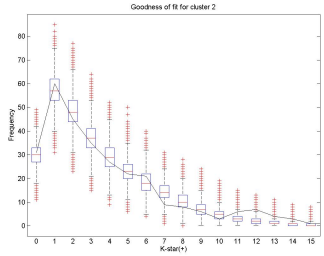
Wilkinson, D. M. and B. A. Huberman (2004). A method for finding communities of related genes. *proceedings of the national Academy of sciences 101*(suppl 1), 5241–5248.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

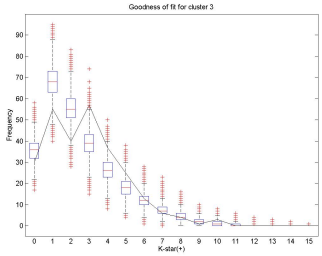
11 Appendix



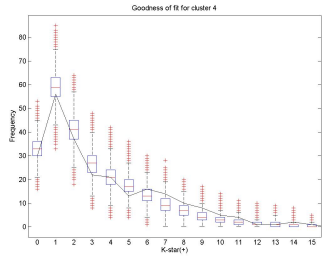
(a) cluster 1



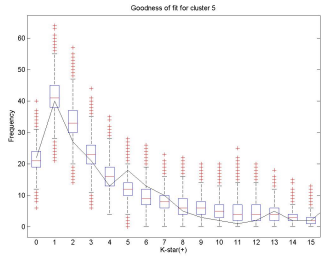
(b) cluster 2



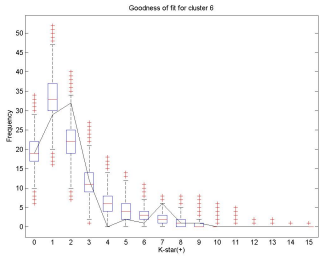
(c) cluster 3



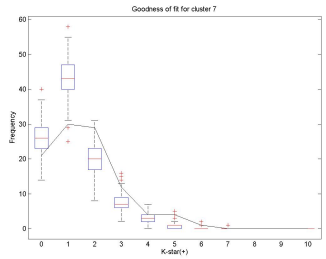
(d) cluster 4



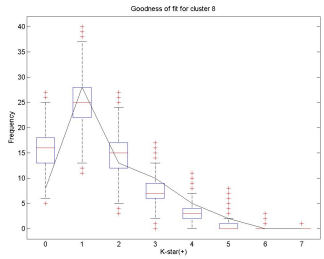
(e) cluster 5



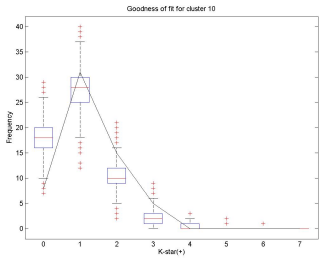
(f) cluster 6



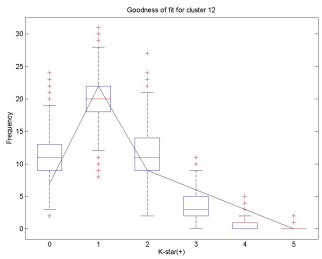
(g) cluster 7



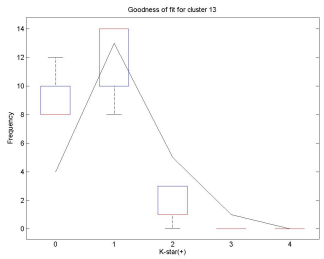
(h) cluster 8



(i) cluster 10

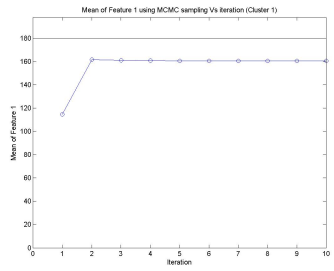


(j) cluster 12

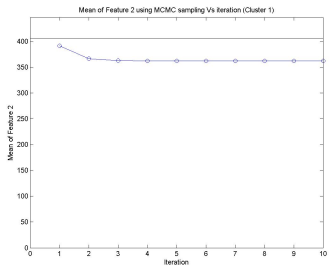


(k) cluster 13

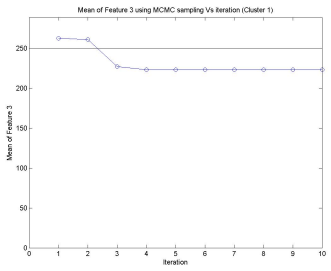
Figure 3: Goodness Of Fit for all cluster (Model 1)



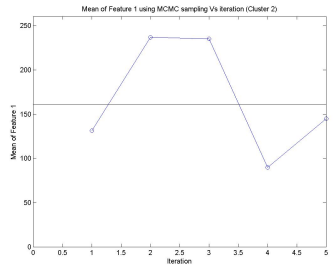
(a) Net. Stat. 1 (cluster 1)



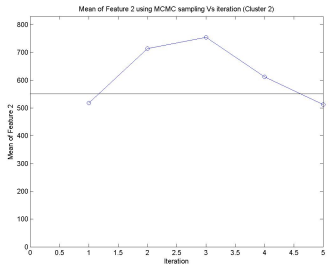
(b) Net. Stat. 2 (cluster 1)



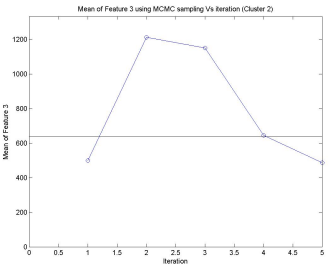
(c) Net. Stat. 3 (cluster 1)



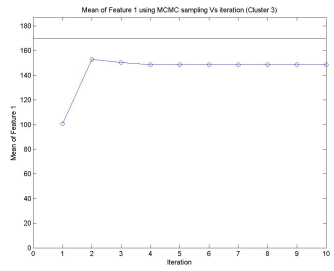
(d) Net. Stat. 1 (cluster 2)



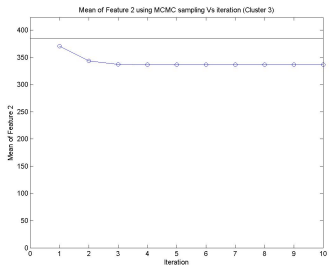
(e) Net. Stat. 2 (cluster 2)



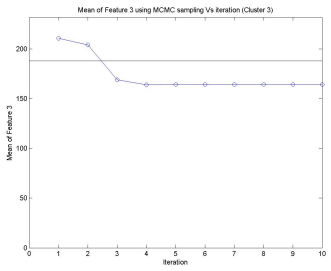
(f) Net. Stat. 3 (cluster 2)



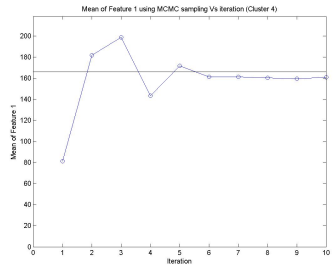
(g) Net. Stat. 1 (cluster 3)



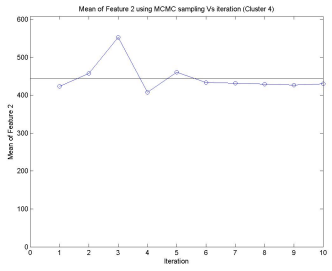
(h) Net. Stat. 2 (cluster 3)



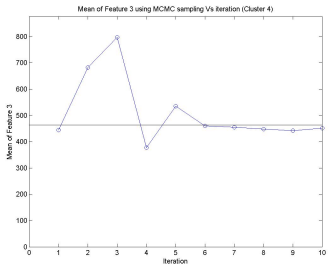
(i) Net. Stat. 3 (cluster 3)



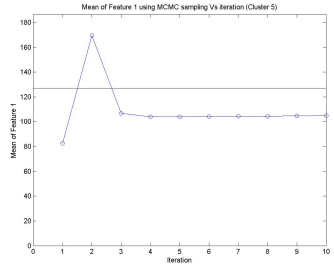
(j) Net. Stat. 1 (cluster 4)



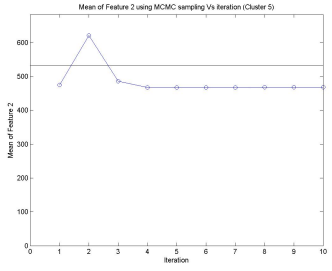
(k) Net. Stat. 2 (cluster 4)



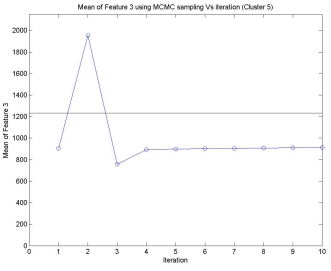
(l) Net. Stat. 3 (cluster 4)



(m) Net. Stat. 1 (cluster 5)

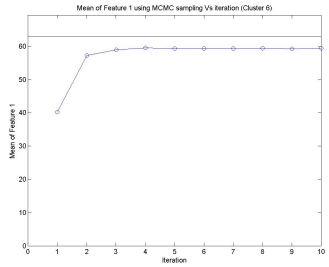


(n) Net. Stat. 2 (cluster 5)

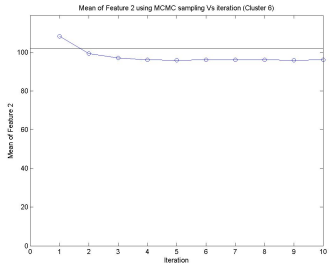


(o) Net. Stat. 3 (cluster 5)

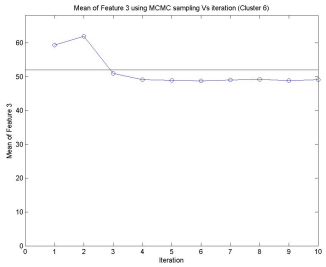
Figure 4: Convergence check in Net. Stat. (Model 1) for cluster 1,2,3,4 & 5



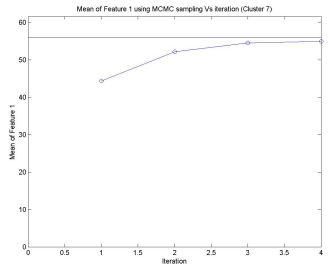
(a) Net. Stat. 1 (cluster 6)



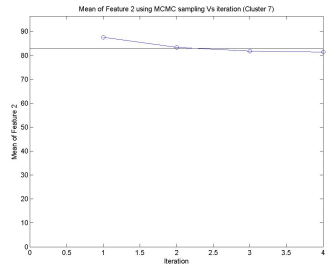
(b) Net. Stat. 2 (cluster 6)



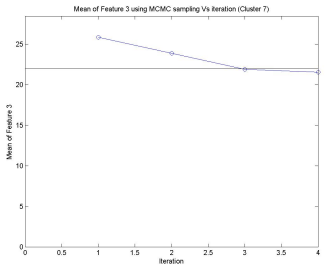
(c) Net. Stat. 3 (cluster 6)



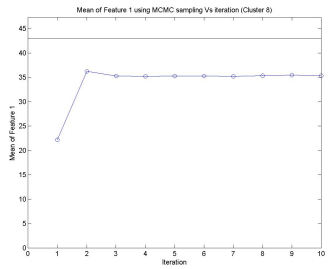
(d) Net. Stat. 1 (cluster 7)



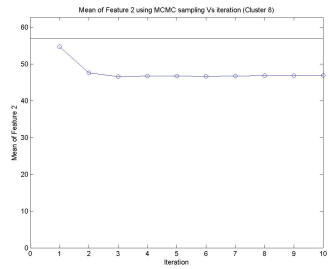
(e) Net. Stat. 2 (cluster 7)



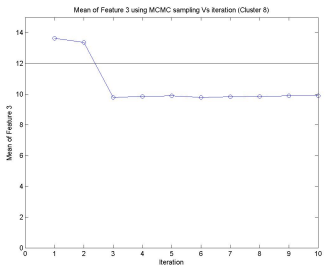
(f) Net. Stat. 3 (cluster 7)



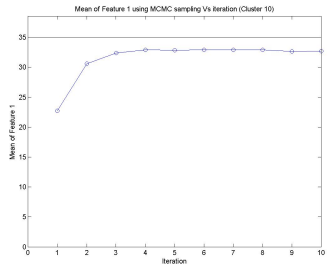
(g) Net. Stat. 1 (cluster 8)



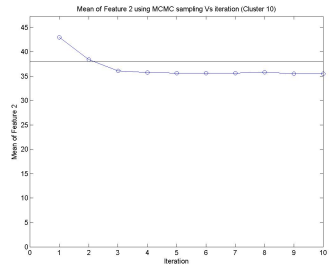
(h) Net. Stat. 2 (cluster 8)



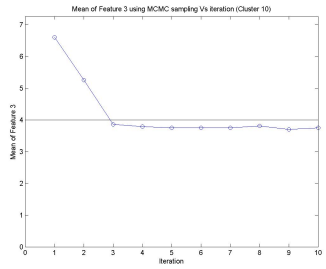
(i) Net. Stat. 3 (cluster 8)



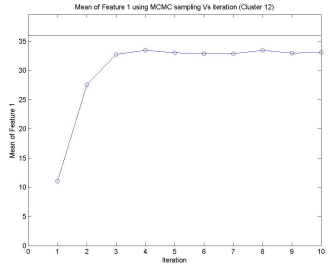
(j) Net. Stat. 1 (cluster 10)



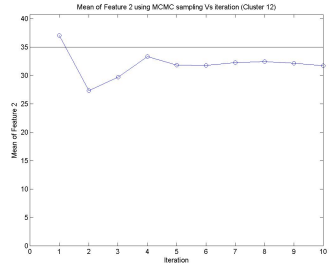
(k) Net. Stat. 2 (cluster 10)



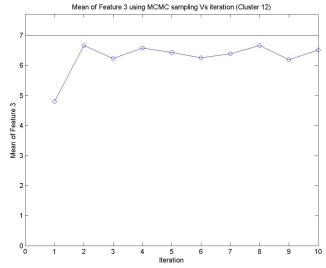
(l) Net. Stat. 3 (cluster 10)



(m) Net. Stat. 1 (cluster 12)

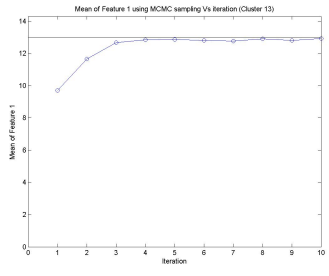


(n) Net. Stat. 2 (cluster 12)

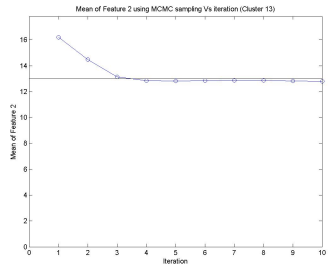


(o) Net. Stat. 3 (cluster 12)

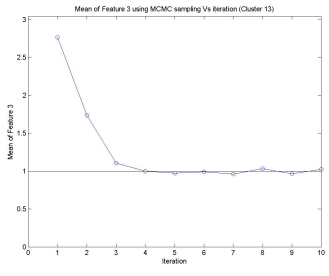
Figure 5: Convergence check in Net. Stat. (Model 1) for cluster 6,7,8,10 & 12



(a) Net. Stat. 1 (cluster 13)

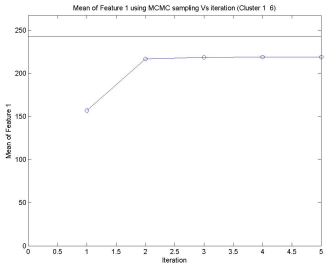


(b) Net. Stat. 2 (cluster 13)

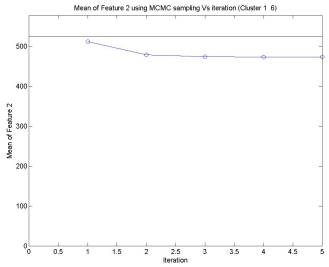


(c) Net. Stat. 3 (cluster 13)

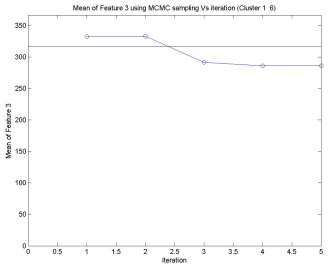
Figure 6: Convergence check in Net. Stat. (Model 1) for cluster 13



(a) Net. Stat. 1 (cluster 1, 6)

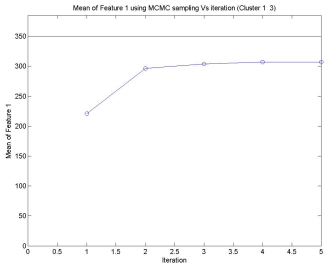


(b) Net. Stat. 2 (cluster 1, 6)

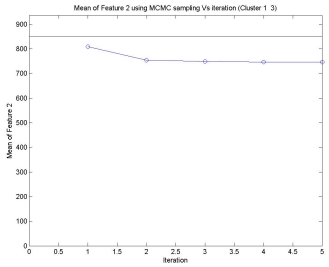


(c) Net. Stat. 3 (cluster 1, 6)

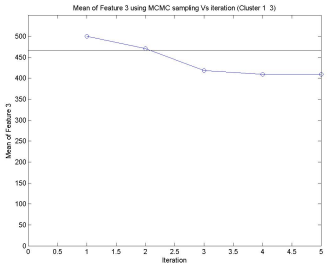
Figure 7: Convergence check in Net. Stat. for combine cluster 1 & 6



(a) Net. Stat. 1 (cluster 1, 3)

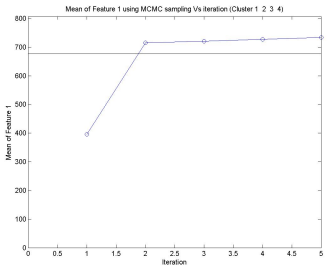


(b) Net. Stat. 2 (cluster 1, 3)

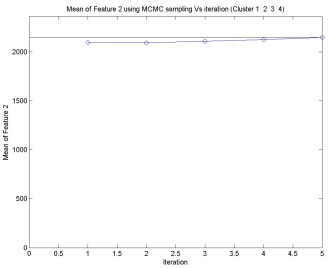


(c) Net. Stat. 3 (cluster 1, 3)

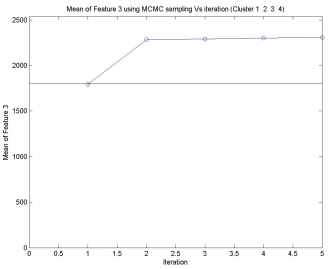
Figure 8: Convergence check in Net. Stat. for combine cluster 1, 3



(a) Net. Stat. 1 (cluster 1,2,3,4)

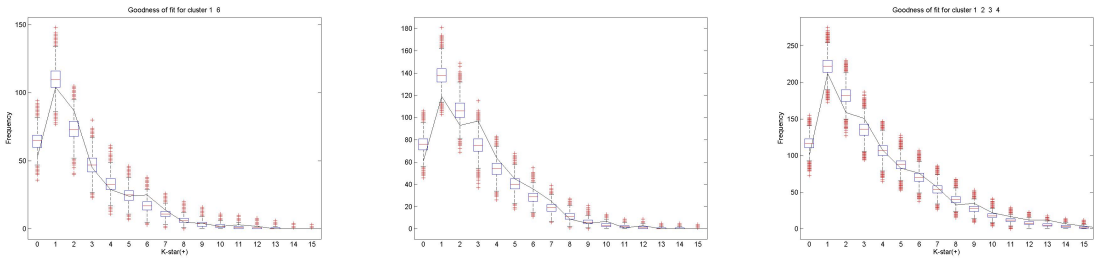


(b) Net. Stat. 2 (cluster 1,2,3,4)



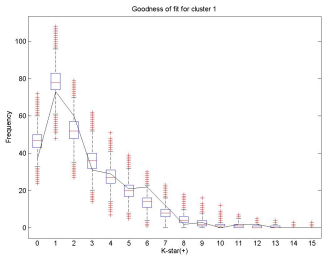
(c) Net. Stat. 3 (cluster 1,2,3,4)

Figure 9: Convergence check in Net. Stat. for combine cluster 1,2,3,4

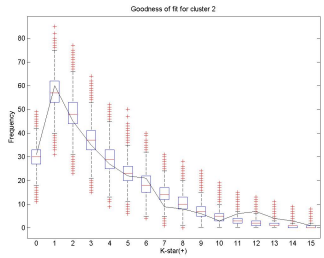


(a) cluster 1 & 6 (b) cluster 1 & 3 (c) cluster 1 & 2 & 3 & 4

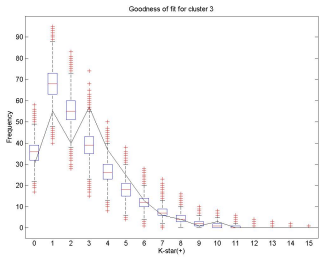
Figure 10: Goodness Of Fit for combine clusters



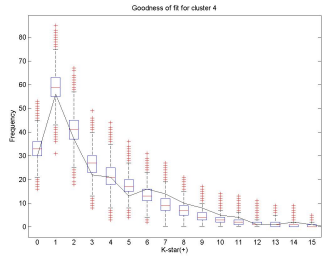
(a) cluster 1



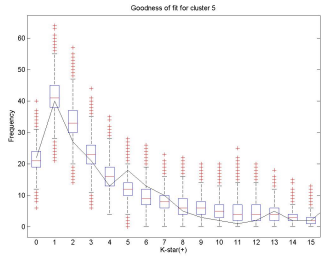
(b) cluster 2



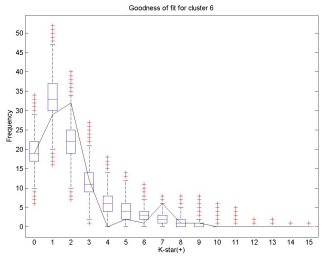
(c) cluster 3



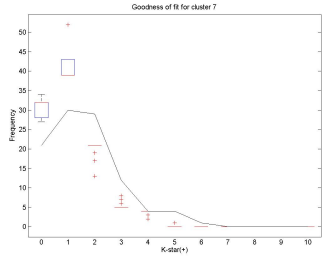
(d) cluster 4



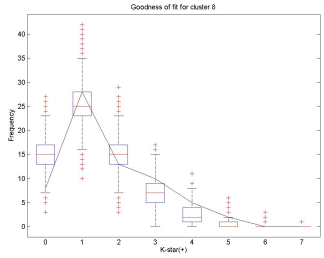
(e) cluster 5



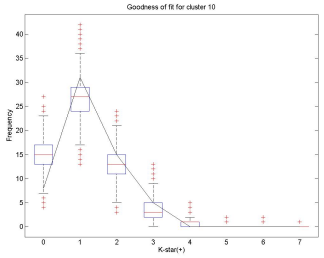
(f) cluster 6



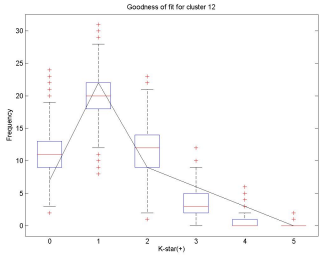
(g) cluster 7



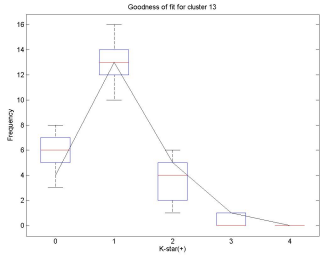
(h) cluster 8



(i) cluster 10

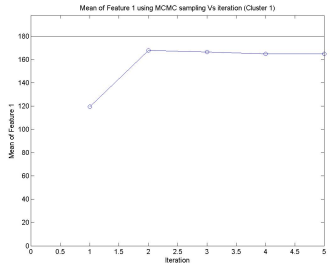


(j) cluster 12

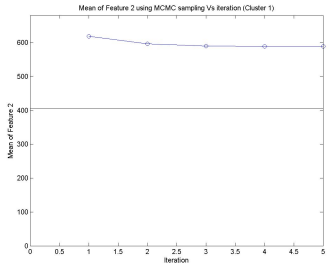


(k) cluster 13

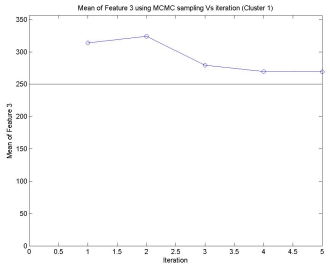
Figure 11: Goodness Of Fit for all cluster (Model 3)



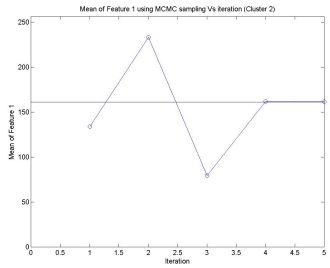
(a) Net. Stat. 1 (cluster 1)



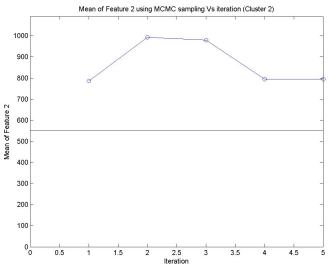
(b) Net. Stat. 2 (cluster 1)



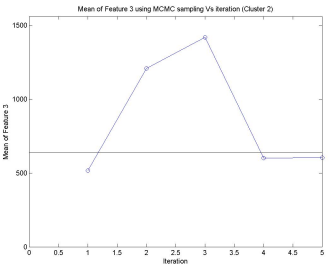
(c) Net. Stat. 3 (cluster 1)



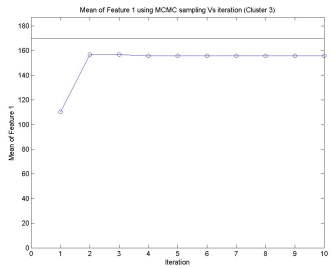
(d) Net. Stat. 1 (cluster 2)



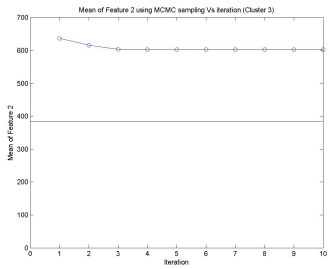
(e) Net. Stat. 2 (cluster 2)



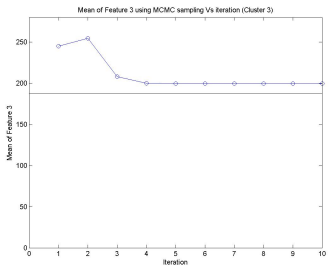
(f) Net. Stat. 3 (cluster 2)



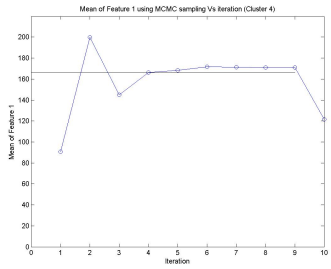
(g) Net. Stat. 1 (cluster 3)



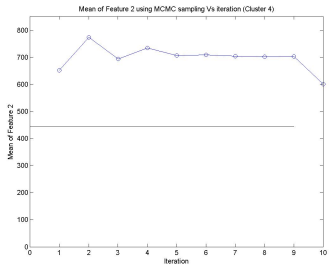
(h) Net. Stat. 2 (cluster 3)



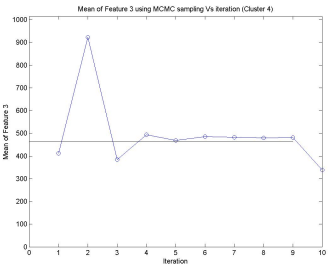
(i) Net. Stat. 3 (cluster 3)



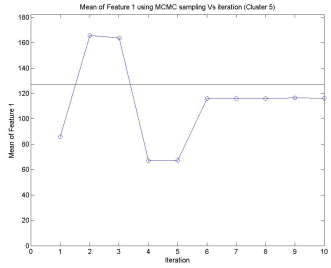
(j) Net. Stat. 1 (cluster 4)



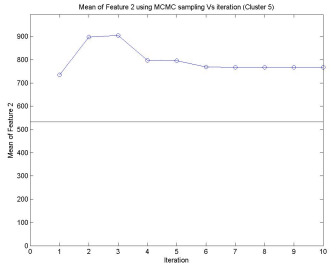
(k) Net. Stat. 2 (cluster 4)



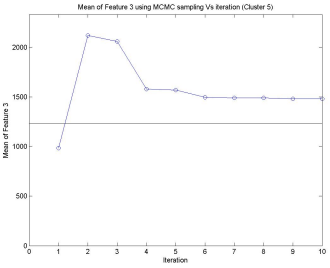
(l) Net. Stat. 3 (cluster 4)



(m) Net. Stat. 1 (cluster 5)

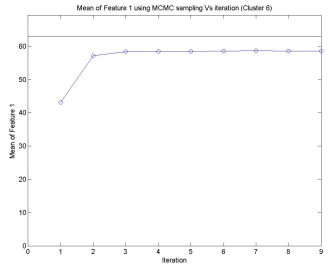


(n) Net. Stat. 2 (cluster 5)

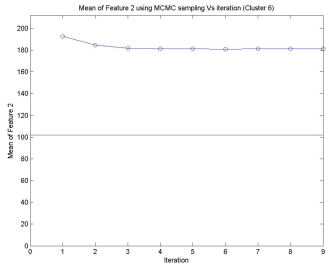


(o) Net. Stat. 3 (cluster 5)

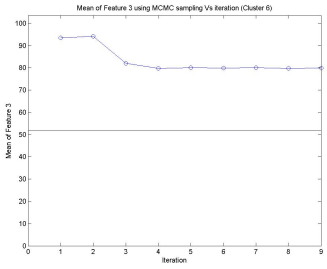
Figure 12: Convergence check in Net. Stat. (Model 3) for cluster 1, 2, 3, 4 & 5



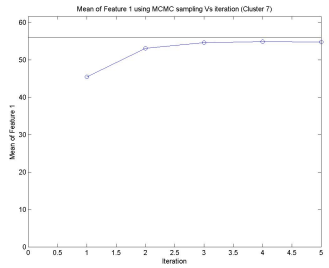
(a) Net. Stat. 1 (cluster 6)



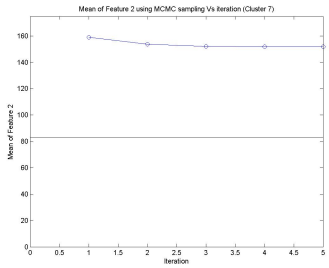
(b) Net. Stat. 2 (cluster 6)



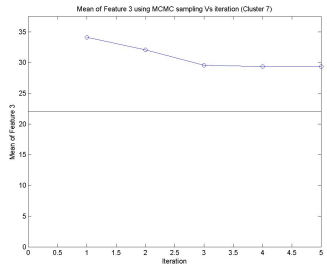
(c) Net. Stat. 3 (cluster 6)



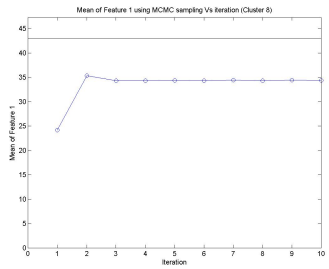
(d) Net. Stat. 1 (cluster 7)



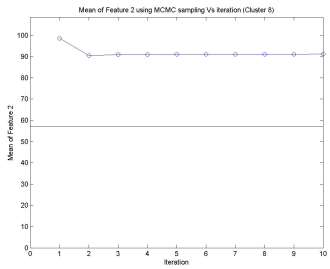
(e) Net. Stat. 2 (cluster 7)



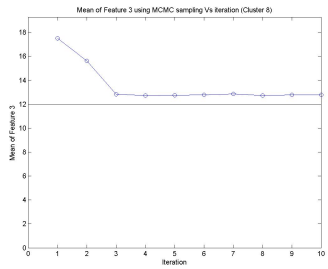
(f) Net. Stat. 3 (cluster 7)



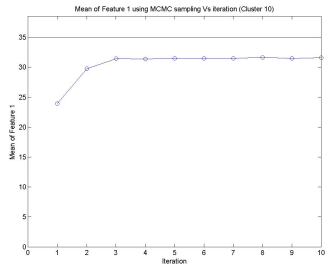
(g) Net. Stat. 1 (cluster 8)



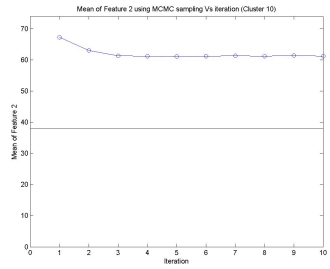
(h) Net. Stat. 2 (cluster 8)



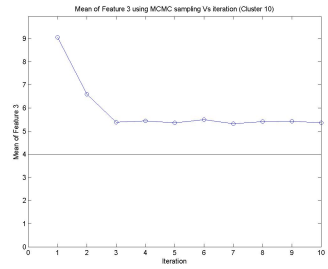
(i) Net. Stat. 3 (cluster 8)



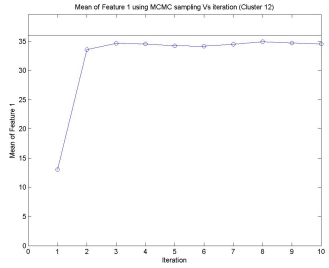
(j) Net. Stat. 1 (cluster 10)



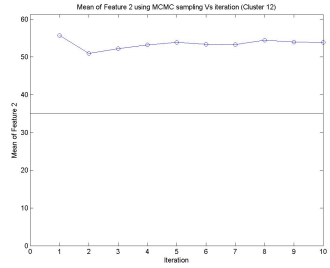
(k) Net. Stat. 2 (cluster 10)



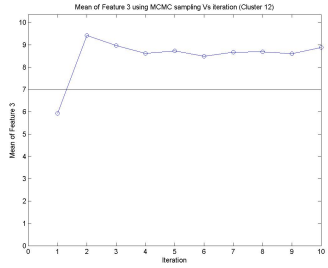
(l) Net. Stat. 3 (cluster 10)



(m) Net. Stat. 1 (cluster 12)

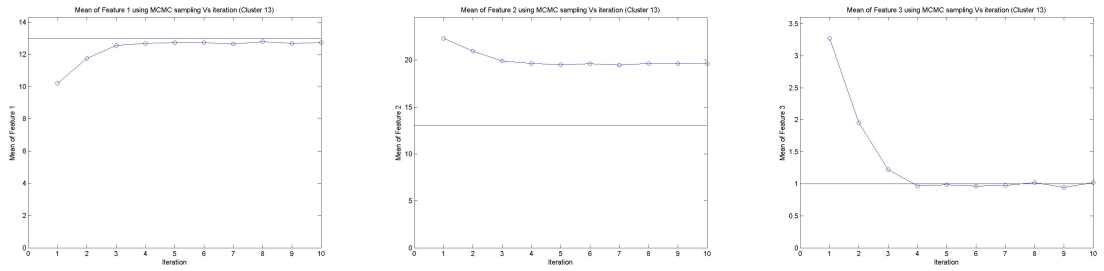


(n) Net. Stat. 2 (cluster 12)



(o) Net. Stat. 3 (cluster 12)

Figure 13: Convergence check in Net. Stat. (Model 3) for cluster 6, 7, 8, 10 & 12



(a) Net. Stat. 1 (cluster 13) (b) Net. Stat. 2 (cluster 13) (c) Net. Stat. 3 (cluster 13)

Figure 14: Convergence check in Net. Stat. (Model 3) for cluster 13