Privacy-Constrained Network Formation [*]

Daron Acemoglu [†] Ali Makhdoumi [‡] Azarakhsh Malekian [§]
Asuman Ozdaglar [¶]

Abstract

With the increasing ease with which information can be shared in social media, the issue of privacy has become central for the functioning of various online platforms. In this paper, we consider how privacy concerns affect individual choices in the context of a network formation game (where links can be interpreted as friendships in a social network, connections over a social media platform or trading activity in online platform). In the model, each individual decides which other agents to "befriend", i.e., form links with. Such links bring direct (heterogeneous) benefits from friendship and also lead to the sharing of information. But such information can travel over other linkages (e.g., shared by the party acquiring the information with others) through a percolation process over the equilibrium network. Privacy concerns are modeled as a disutility that individual suffers as a result of her private information being acquired by others, and imply that the individual has to take into account who the friends of her new friend (and who the friends of friends of her new friend etc.) are. We specify conditions under which pure-strategy equilibria exist and characterize both pure-strategy and mixed-strategy equilibria. Our two main results show that, as in many real-life examples, the resulting equilibrium networks feature clustered connections and homophily. Clustering emerges because if player a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, thus making it less costly for a to befriend c. Homophily emerges because even an infinitesimal advantage in terms of direct benefits of friendship within a group makes linkages within that group more likely, and the travel of information within that group reduces the costs, and thus increases the likelihood, of further within-group links.

Keywords— Network formation, Networks, Privacy.
JEL Classification : D85

[*]We thank various numerous seminar and conference participants for useful suggestions. We gratefully acknowledge financial support from the Toulouse Network with Information Technology and Army Research Office, ARO MURI W911NF-12-1-0509.

[†]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, 02139 . daron@mit.edu

[‡]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 makhdoum@mit.edu

[§]Rotman School of Management, University of Toronto, Toronto, ON, M5S 3E6 azarakhsh.malekian@rotman.utoronto.ca

[¶]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 asuman@mit.edu

1 Introduction

With the increasing volume of information-sharing in online platforms, privacy has become a central concern. Many individuals are willing to take costly actions in order to prevent platforms, retailers, advertisers as well as acquaintances from having access to their private information (Varian (1997)). Though many online platforms, including Facebook, have taken steps to alleviate these concerns, privacy is likely to become even more important issue in the years to come.

There is relatively little work, however, on how privacy concerns impact online behavior. In this paper, we take a first step by considering a network formation game in the presence of information leakage over the network, which is costly for individuals. The network in question can stand for the friendship or connections network over a social media platform or as an abstract representation of online trading activity. Thus the insights from our analysis should apply both to social media and to online commercial activities.

In our model, each individual decides to form directed links to others. Links have heterogeneous benefits (e.g., an individual receives benefits from befriending others in a social media platform). Once formed, these links also transmit information, however. For example, a friendship link over a social media platform inevitably involves some information sharing, while online trades will necessarily give information to the user's trading partner about his or her preferences. More important than the direct transmission of this information is indirect transmission over the network: the relevant information can travel not only to one's friend but also to friends of friends, etc. This makes the cost of loss of privacy for an individual a function of the equilibrium network (which other friendships have formed in equilibrium).

Though the social interactions captured by our model are potentially complex, the setup is relatively parsimonious. It consists of a matrix of benefits of direct links, a cost of loss of privacy (the cost of an individual's information being observed by each other agent in the network), and percolation process for the travel of information over the network.

We first characterize properties of the equilibrium network. Our first result identifies an endemic problem of non-existence of pure-strategy equilibria. This can be best understood by considering relationship between three agents. If player b has formed a link to player c, this will discourage player a from forming a link with player b, because any information shared with b now risks travelling to c. This in turn, encourage player c to form a link with player a. Finally, since c has formed a link with a, b would disconnect from c, resulting in a contradiction.

We also establish that a sufficient and necessary condition for the existence of pure-strategy equilibria is that the matrix of benefits of direct links is such that high benefit subset of edges have no cycle. In particular, we show that if this condition holds, then a pure-strategy Nash equilibrium always exists and if this condition does not hold, there exist a set of popularities for which no pure-strategy Nash equilibrium exists.

As a final characterization result, we also establish an interesting phase transition in pure-strategy equilibria as we vary the ability of information transmission. As this probability increases, there are two opposing forces: first, with higher transmission probability, an agent's information is likely to reach any other agent she is indirectly connected to, and this discourages connections. But secondly, and in contradiction to the first force, this greater transmission probability also implies that the cost of connecting directly to such an agent is lower, thus encouraging greater connections. We show that the resolution of these two opposing forces implies that until a critical value of this threshold is reached, the equilibrium network is sparse, but as this critical value is reached, the equilibrium becomes a collection of densely-connected cliques with clustering coefficient one.

Our other sets of results concern the patterns of connections that occur (in pure or mixed-strategy equilibria). Our first major result shows that equilibrium networks feature clustered connections. This pattern emerges because if a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, thus making it less costly for a to befriend c. Second, we also show that the equilibrium network features homophily. The reason for this is that even an infinitesimal advantage in terms of direct benefits of friendship within a group makes linkages within that group more likely, in turn making information travel within that group and reducing the cost of making further within-group links due to loss of privacy. This increases the likelihood of further within-group links.

Though there is relatively little work on how privacy affects individual decisions in online platforms and social media settings, there are several other large and growing literature to which our paper relates.[1] First, our work is part of a large literature on endogenous social networks. Key works here include Tardos and Wexler (2007), Barabási and Albert (1999), Chung and Lu (2002a), Chung and Lu (2002b), Jackson and Rogers (2007), Newman (2003), Newman (2004), Watts and Strogatz (1998), Galeotti et al. (2006), Skyrms and Pemantle (2009), Blume et al. (2011). Perhaps more closely related is Currarini

[1]See Dwork and Roth (2013) and Liang et al. (2009) for surveys of various aspects of privacy concerns. Calzolari and Pavan (2001), Laudon (1996), Varian (1997), Taylor (2004), and Hui and Png (2006) for examples of economics papers discussing privacy-related issues, and Samuelson (2000), Westin (1967), Stigler (1980), Hirshleifer (1980), and Magi (2011) for certain legal aspects of privacy.

et al. (2009), which develops a model of friendship formation where individulas receive type-dependent benefits from friendship, and explains the emergence of homophily in friendships and how this varies with group size.

Second, a large sociology and network literature emphasizes the regularity of triadic closure. In the words of Rapoport (1953): *"If two people in a social network have a friend in common, then there is an increased likelihood that they become friends themselves at some point in the future."* This pattern can be detected either by verifying triadic closure properties or focusing on various network statistics that provide information on this, such as the *clustering coefficient*, which measures the probability that two randomly selected friends of a node are friends with each other (e.g., Newman (2003), Watts and Strogatz (1998) Fagiolo (2007)). Evidence on these patterns is provided in, among others, Medus and Dorso (2013), Kossinets and Watts (2006), Albert and Barabási (2002), Davidsen et al. (2002), Holme and Kim (2002), and Vázquez (2003).

Third, there is also a similarly large literature on the second key pattern generated by privacy-constrained network formation: homophily. Homophily is defined as the tendency of people to associate with others similar to themselves, is observed in many social networks, ranging from friendships to marriages to business relationships, and is based on a variety of characteristics and attributes, including ethnicity (see Fong and Isajiw (2000) and Baerveldt et al. (2004) for examples of studies focusing on ethnicity), race, age, gender, religion, education level, profession, political affiliation , and other attributes (see for example Lazarsfeld et al. (1954), Blau (1977), Blalock (1982), Marsden (1988), Marsden (1987), or the survey by McPherson et al. (2001)). Various different explanations for homophily have been proposed in, among others, Moody (2001), Patacchini and Zenou (2006), Currarini et al. (2009), and Fowler et al. (2009).

Also closely related are Kleinberg et al. (2001); Blume et al. (2011); Fabrikant et al. (2003) who study a model of the trade-off between the benefits received from sharing information and the cost of indirect sharing of information, though both their models and results are very different from ours.

2 Model

We consider a set $\mathcal{V} = \{1, \ldots, n\}$ of agents interested in forming friendship links with each other. In choosing their links, agents tradeoff the benefit from direct links with the cost of privacy loss due to leakage of information through indirect links. Each agent makes a decision about connecting to other agents, i.e., agent i chooses a strategy $\mathbf{x}_i =$

$(x_{i1}, \ldots, x_{in})$, where $x_{ij} \in \{0, 1\}$ represents whether agent i is connected to agent j or not (we use the convention that $x_{ii} = 0$ for all $i \in \mathcal{V}$). We assume that the decision $x_{ij} = 1$ results in a directed friendship link from i to j, implying i shares her information with j and receives friendship benefits from it but not necessarily the other way around, i.e., x_{ji} need not be equal to x_{ij}. This means that if agent i shares her information with j (e.g., to get advise on a matter), agent j does not have to share her information with i. Though most prior literature (e.g., see Jackson (2005)) considers undirected models of friendship in the context of network formation, in several settings links and friendships are not always on equal footing and can be more fruitfully modeled as directed links. For instance, a friendship might involve one party, individual a, sharing information with another, b, either in the course of social interactions or to receive some advice, while b does not share any information in return. In the context of social media, the amount of information shared between friends and connected individuals is again often asymmetric. Finally, another application of these ideas would be to other online interactions, such as individuals using a website, platform or service, and uploading information in the process (e.g., likes and dislikes or credit card information).[2]

Given node i's decision x_{ij} with respect to agent j, agent i derives a benefit $v_{ij}x_{ij}$ from her friendship with agent j, where $v_{ij} \geq 0$ is a parameter that captures the value i has for her friendship with j. We collect all v_{ij}'s in an $n \times n$ matrix $\mathbf{V} = [v_{ij}]_{i,j \in \mathcal{V}}$ and refer to it as the *valuation matrix* (we use the convention that $v_{ii} = 0$ for all $i \in \mathcal{V}$). We also collect the strategy of all agents in a $n \times n$ matrix $\mathbf{x}$ where $[\mathbf{x}]_{ij} = x_{ij}$. The matrix $\mathbf{x}$ is the adjacency matrix of the formed graph by the strategies of agents. Finally, we let $\mathbf{x}_{-i} = (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$ to show the strategy profile of all agents other than i. Next, we formally define the graph (network) formed given the strategy of agents, the process for the indirect leakage of information, and the utility function of the agents.

A given strategy profile $\mathbf{x}$ induces a (directed) graph among agents $G_{\mathbf{x}} = (\mathcal{V}, E_{\mathbf{x}})$, where $E_{\mathbf{x}}$ is the set of directed edges given by

$$E_{\mathbf{x}} = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid x_{ij} = 1\},$$

where $(i, j) \in E_{\mathbf{x}}$ implies that in the graph $G_{\mathbf{x}}$ there is a direct link from i to j. Information from agent (node) i leaks to other agents over $G_{\mathbf{x}}$ according to the following probabilistic process:

[2] This last application would require other, relatively straightforward, changes (e.g., considering a bipartite graph users and online platforms, with information flows in only one direction, and making decisions about which platforms are linked and share information among themselves).

- Information from i reaches each of her out-neighbors $l \in N^{\text{out}}(i)$ with probability 1, where $N^{\text{out}}(i) = \{l \in \mathcal{V} \ : \ (i,l) \in E_{\mathbf{x}}\}$.
- Any information that agent l has, e.g. about agent i, she sends it to any of her out-neighbors with probability β independent of everything else.

We refer to $\beta \in [0,1]$ as the *transmission parameter*. This process essentially assumes that information from i leaks according to the independent cascade model (see Kempe et al. (2003)) with transmission probability 1 on the neighboring edges and transmission probability β on all other edges. It captures the assumption that direct neighbors/friends of i have access to all information of i, while indirect friends may obtain information of i through a probabilistic gossip process.

For a given node i, in order to compute the probability that information from i reaches node j, denoted by $\mathbb{P}[i \rightsquigarrow j]$, we consider another equivalent view of the information leakage process. We draw a realized graph $G = (\mathcal{V}, E)$, on which we activate outgoing edges of i on $G_{\mathbf{x}}$ with probability 1 and all other edges of $G_{\mathbf{x}}$ with probability β (therefore, we have $E \subseteq E_{\mathbf{x}}$). Information from i will reach node j if and only if there is a directed path of active edges from i to j on this realized graph, in which case we say node j is reachable from node i and write $i \rightsquigarrow j$.[3] The probability that the graph G is realized is

$$P_i(G) = \left(\prod_{(i,l) \in E_{\mathbf{x}}} \mathbf{1}\{(i,l) \in E\} \right) \left(\prod_{\substack{(k,k') \in E_{\mathbf{x}},\ k \neq i \\ (k,k') \in E}} \beta \right) \left(\prod_{\substack{(k,k') \in E_X,\ k \neq i \\ (k,k') \notin E}} (1-\beta) \right),$$

where the subscript i denotes the dependence of this probability on agent i and $\mathbf{1}\{.\}$ is the indicator function.

Therefore, the probability that information from i reaches node j is given by

$$\mathbb{P}[i \rightsquigarrow j] = \sum_{G \in \mathcal{G}} P_i(G) \mathbf{1}\{i \rightsquigarrow j\},$$

where $\mathcal{G}$ denotes the set of all possible graphs with the set of nodes $\mathcal{V}$ and the set of edges which is a subset of E_X, i.e.,

$$\mathcal{G} = \{G = (\mathcal{V}, E) \mid E \subseteq E_{\mathbf{x}}\}.$$

The aggregate leakage of information of i, denoted by $\text{Gossip}(i, \mathbf{x})$ is the summation of probabilities $\mathbb{P}[i \rightsquigarrow j]$ over all $j \in \mathcal{V}$. With the convention that for any $i \in \mathcal{V}$, $\mathbb{P}[i \rightsquigarrow i] = 0$,

[3] Throughout the paper, we use the notation $i \rightsquigarrow j$ to denote that j is reachable from i, and the notation $i \to j$ to show that i has a direct link to j, i.e., $x_{ij} = 1$. We will also use the notation $i \not\to j$ to denote $x_{ij} = 0$.

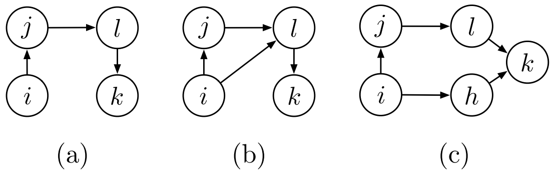


Figure 1: An example that illustrates the calculation of gossip probability.

it can be written as

$$\text{Gossip}(i, \mathbf{x}) = \sum_{j \in \mathcal{V}} \mathbb{P}[i \rightsquigarrow j].$$

Note that $\text{Gossip}(i, \mathbf{x})$ is a function of the adjacency matrix $\mathbf{x}$ as well as the transmission parameter β.

For instance, consider the networks given in Figure 1a. The probability that the information from i reaches k is equivalent to having both edges (j, l) and (l, k) active (edge (i, j) is active with probability one), which happens with probability β^2, therefore, $\mathbb{P}[i \rightsquigarrow k] = \beta^2$. For the network given in Figure 1b, since i is connected to l and j, they both have the information of i with probability one and i's information reaches k if (l, k) is active, implying $\mathbb{P}[i \rightsquigarrow k] = \beta$. Finally, for the network given in Figure 1c, the probability that the information from i reaches k is to either have the edges (j, l) and (l, k) being active, or to have edge (h, k) being active, or both events, therefore, $\mathbb{P}[i \rightsquigarrow k] = \beta^2(1 - \beta) + \beta(1 - \beta^2) + \beta^3 = \beta^2 + \beta - \beta^3$.

The utility function of agent i, denoted by u_i, is given by

$$u_i(X) = \sum_{j \in \mathcal{V}} x_{ij} v_{ij} - \gamma \sum_{j \in \mathcal{V}} \mathbb{P}[i \rightsquigarrow j], \tag{2.1}$$

where the *cost parameter* $\gamma \geq 0$ captures the tradeoff between value of friendship and loss of privacy. For a given valuation matrix $\mathbf{V}$, which is assumed to be known by all agents, we define the Nash equilibrium of the complete information game as follows.

Definition 1 (Pure-Strategy Nash Equilibrium) The set of strategies $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a pure-strategy Nash equilibrium if for all $i \in \mathcal{V}$, we have

$$\mathbf{x}_i \in \text{argmax}_{\mathbf{y_i} \in \{0,1\}^n} u_i(\mathbf{y_i}, \mathbf{x}_{-i}).$$

We refer to the network induced by the strategies $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as the *equilibrium network*, where $\mathbf{x}$ shows its adjacency matrix.

We focus on Nash equilibria, rather than pairwise stability as in Jackson and Wolinsky (1996) (see also Jackson (2005) for an overview of other solution concepts), since we wish to focus on each individuals' incentive to form links unilaterally in the context of a directed friendship network.[4]

Because, as we show below, pure strategy Nash equilibria may not always exist, we also consider mixed-strategy Nash equilibria, defined in the usual fashion.

Definition 2 (Mixed-Strategy Nash Equilibrium) *The mixed strategy* $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$*, where* σ_i *is a probability measure over* $\{0,1\}^n$*, is a mixed Nash equilibrium if for any* i*, we have*

$$u_i(\sigma_i, \boldsymbol{\sigma}_{-i}) \geq u_i(\mathbf{y}_i, \sigma_{-i}), \text{ for any } \mathbf{y}_i \in \{0,1\}^n,$$

where $\boldsymbol{\sigma}_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$*.*

Finally, when we turn to the analysis of homophily, there will sometimes be additional, unintuitive equilibria. One way of eliminating these is to consider strong (pure-strategy) Nash equilibria, which also test for deviations by coalitions.(We will also establish that our other results are valid regardless of whether we use pure-strategy Nash equilibrium or strong Nash equilibrium).

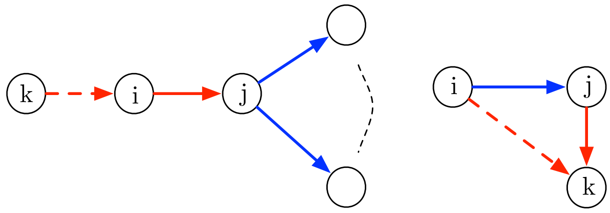
Definition 3 (Strong Nash Equilibrium) A set of decisions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a strong pure-strategy Nash equilibrium if and only if there exists no *coalition* $S \subseteq \{1, \ldots, n\}$ that has a profitable deviation, i.e., none of the agents of S receives a lower utility after deviation and at least one of the agents of S receives a higher utility.

The analysis of the Nash equilibrium of this game is made complicated by the fact that it features both strategic complementarity and strategic substitutability. In particular, a link from an agent i to agent j generally discourages another agent k from forming a link to i because of increased likelihood of leakage of k's information, introducing an element of strategic substitutes (see Figure 2a). On the other hand, when there is already a link from i to j, then if j forms a link to k, i would become more likely to form a link to k because her information is already leaked indirectly to k, which is an element of strategic complements (see Figure 2b).

3 Existence of Equilibria

In this section, we study the existence of both pure and mixed-strategy Nash equilibria. When either transmission parameter β or cost parameter γ is zero, a pure-strategy Nash

[4]This is particularly natural in the case of connections in the context of social media, which are unilateral, directed links. In the case of friendship, we interpret links to represent how much trust an individual puts in an acquaintance or member of broader community, which is again better represented as a directed link decided unilaterally. In other settings, however, joint decisions might be important because a friendship may require participation by both parties.



(a) (b)

Figure 2: (a): strategic substitutability, (b): strategic complementarity.

equilibrium always exists.[5] In the rest of this section, we will focus on strictly positive values of transmission parameter $\beta > 0$ and cost parameter $\gamma > 0$.

We first give an example in which a pure-strategy Nash equilibrium does not exist.

Example 1 Let $\gamma > 0$ and $\beta > 0$. Consider three agents, denoted by a, b, c, and let $v_{ab} = v_{bc} = v_{ca} = \gamma(1 + \frac{\beta}{4})$ and $v_{ba} = v_{cb} = v_{ac} = \gamma(1 - \frac{\beta}{2})$. First note that in any pure-strategy Nash equilibrium, we have $x_{ba} = x_{cb} = x_{ac} = 0$. To see this suppose the contrary that $x_{ba} = 1$. Therefore, the utility of b is upper bounded by

$$u_b(\mathbf{x}) \leq \max\{v_{ba} - \gamma, v_{ba} + v_{bc} - 2\gamma\} < 0,$$

which is negative and a profitable deviation for b would be to disconnect from both a and c in order to obtain zero utility. Therefore, an equilibrium network (if exists) belongs to the set of eight possible networks defined by $(x_{ab}, x_{bc}, x_{ca}) \in \{0, 1\}^3$. Next, we will argue that none of them can be an equilibrium network.

- an empty network is not an equilibrium network as a has a profitable deviation which is $x_{ab} = 1$. This deviation would increase her utility from 0 to $v_{ab} - \gamma = \gamma\frac{\beta}{4}$.
- a network with one edge in not an equilibrium network. Without loss of generality (because of symmetry), suppose $x_{ab} = 1$. Player b has a profitable deviation which is $x_{bc} = 1$. This deviation would increase her utility from 0 to $v_{bc} - \gamma = \gamma\frac{\beta}{4}$.
- a network with two edges is not an equilibrium network. Without loss of generality, suppose $x_{ab} = x_{bc} = 1$. Player a has a profitable deviation, i.e., $x_{ab} = 0$. This deviation would increase her utility from $v_{ab} - \gamma(1 + \beta) = -\frac{3}{4}\gamma\beta$ to 0.

[5] If transmission parameter $\beta = 0$, for any agent i and j where $v_{ij} \geq \gamma$ we let $x_{ij} = 1$. These decisions clearly form an equilibrium. Similarly, if $\gamma = 0$, then a complete graph is a pure Nash equilibrium. The cases $\beta = 0$ or $\gamma = 0$ correspond to situations where agents do not face any loss of utility due to privacy breach and there will be no trade-off between the benefit of friendship and the cost of indirect leakage of information.

- a network with three edges is not an equilibrium network for the same reason as in the previous case.

As suggested in Example 1, if there exists a directed cycle on the set of the edges with valuation higher than the cost γ, then a pure-strategy Nash equilibrium might not exist. We will next establish a necessary and sufficient condition for the existence of a Nash equilibrium in terms of the graph formed by edges whose valuation parameter are "high", denoted by popular-connections which is defined next.

Definition 4 (Popular-Connections Graph) For a given valuation matrix $\mathbf{V}$, a connection (i, j) is called *popular* if it has valuation at least γ, i.e., $v_{ij} \geq \gamma$. We then define the popular-connections graph as a directed graph with vertex set $\mathcal{V}$ with an edge between two nodes i and j if and only if (i, j) is popular, i.e., popular-connections graph is the graph $(\mathcal{V}, E_\gamma)$ where

$$E_\gamma = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid v_{ij} \geq \gamma\}.$$

In words, the popular-connections graph includes edges where the direct benefit of connection, v_{ij}, exceeds the direct cost from loss of privacy, γ. It is also useful to observe that any edge in this graph would be formed if there were no other edges formed (since in that case the total cost of loss of privacy is exactly γ).

In the next theorem, we show that the necessary and sufficient condition for the existence of a pure Nash equilibrium is indeed the absence of cycles in popular-connections graph.

Theorem 1 *Let $\beta > 0$ and $\gamma > 0$. Sufficient and necessary conditions for the existence of pure-strategy Nash equilibrium are as follows.*

1. *If the popular-connections graph $(\mathcal{V}, E_\gamma)$ contains a simple cycle (of length at least three),[6] then there exists an assignment of valuation matrix $\mathbf{V}$ and a transmission parameter $\bar{\beta}$ such that for all $\beta < \bar{\beta}$ no pure-strategy Nash equilibrium exist.*

2. *If the popular-connections graph $(\mathcal{V}, E_\gamma)$ has no cycle, then there exists a pure-strategy Nash equilibrium. Furthermore, starting from an empty graph ($\mathbf{x}_i = 0$, for all $i \in \mathcal{V}$), the best response dynamics converges to a pure-strategy Nash equilibrium.*

The proof idea of Theorem 1 is as follows. For the first part, similar to Example 1, we show that if the popular-connections graph has a cycle, then there exists a popularity

[6] A simple cycle of a graph is a cycle with no repetition of nodes.

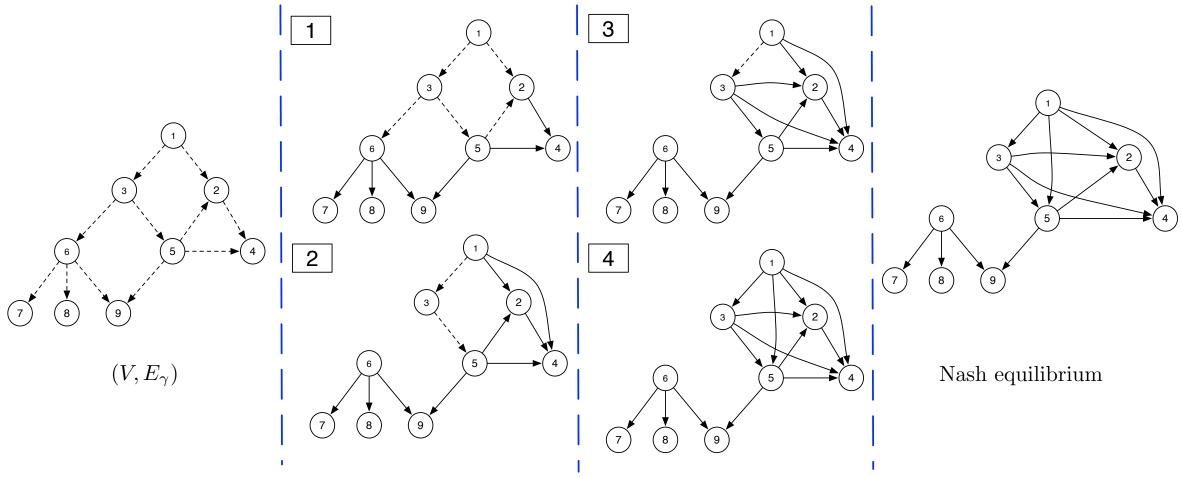


Figure 3: Construction of pure Nash equilibrium

matrix **V** for which no pure strategy Nash equilibrium exists. The proof of the second part is constructive. The idea is to consider the "topological sort" of the agents, which sorts the vertices (agents) of the popular-connections graph in a way that if $a \to b$ in popular-connections graph, then a has a higher rank than b. Because the popular-connections graph is acyclic, such an ordering always exists (see e.g. Leiserson et al. (2001)). We then consider the best response of each agent to the strategies of agents with lower rank, and show that this profile of best responses constitutes a pure strategy Nash equilibrium. This construction is further illustrated in the following example.

Example 2 We will consider $(\mathcal{V}, E_\gamma)$ in Figure 3, where we have 9 nodes. In this example, we suppose that all the popularities along the edges of $(\mathcal{V}, E_\gamma)$ are H, all the other popularities are L, and $\beta = 1$, where $H-\gamma > 0$, $L-\gamma < 0$, and $(H-\gamma)+2(L-\gamma) > 0$, but $(H-\gamma)+3(L-\gamma) < 0$ (this is guaranteed for instance for $H = \frac{9}{4}\gamma$ and $L = \frac{1}{2}\gamma$). This implies that a connection of a to b, where $v_{ab} = H$, can compensate for two connections with L popularities, but not three connections with L popularities. Since popular-connections graph does not have a cycle, we have some nodes with only incoming edges (otherwise, if all nodes have both incoming and outgoing edges, we would have a cycle). In this example those nodes are $\{7, 8, 9, 4\}$. We define a set $R = \{7, 8, 9, 4\}$ and we will update this set in each step of the construction. Each step of the construction has two stages. In the first stage, we let all nodes to play their optimal decision regarding the nodes in R and in the second stage we update the set R.

In the first step, we let all nodes play their optimal decisions regarding nodes in R (given the current set of decisions by others). Here, the decisions are $x_{67} = x_{68} = x_{69} = x_{59} =$

$x_{24} = x_{54} = 1$ (this is labeled by 1 in Figure 3). We then consider all nodes whose out-neighbors are a subset of R. Here, it would be nodes 6 and 2. We then update the set R by including 2 and 6, i.e., $R = \{7, 8, 9, 4, 2, 6\}$. In the second step, we let all players to play their optimal decision regarding the nodes in R. Here, the decisions are $x_{36} = 0$ (since a high value connection does not compensate for three low valuation connections) and $x_{52} = x_{12} = x_{14} = 1$ (this is labeled by 2 in Figure 3). We again update the set R that becomes $R = \{7, 8, 9, 4, 2, 6, 5\}$. In the third step, we let all the players to play their optimal decisions regarding nodes in R that would be $x_{35} = x_{34} = x_{32} = 1$ (this is labeled by 3 in Figure 3). We then update the set R as $R = \{7, 8, 9, 4, 2, 6, 5, 3\}$. In the fourth step, the optimal decisions regarding nodes in R would be $x_{13} = x_{15} = 1$ (this is labeled by 4 in Figure 3), we then update R which would be the entire set of nodes $\mathcal{V}$ and the construction stops. We show that after these steps, the resulting network is an equilibrium network, as shown in the column at the right hand side of Figure 3.

We next show that the pure-strategy Nash equilibrium we have just characterized is also a strong Nash equilibrium, and thus it is robust against deviations by coalitions.

Proposition 1 *Let $\beta > 0$ and $\gamma > 0$. If the popular-connections graph has no cycle (of length at least three), then the constructed pure-strategy Nash equilibrium in Theorem 1 is also a strong Nash equilibrium.*

This result also follows by considering the aforementioned topological sort of the agents. Because, as discussed above, the pure-strategy Nash equilibrium is characterized by the best response of each agent to the strategies of those with lower rank in the topological sort, there is no coalitional deviation that would simultaneously make any subsets of agents better off.

The final remark in this section is that a mixed-strategy Nash equilibrium always exists, regardless of whether he popular-connections graph has a cycle, as shown in the next proposition.

Proposition 2 *Given any valuation matrix $\mathbf{V}$ and transmission parameter $\beta \in [0, 1]$ there always exists a mixed Nash equilibrium.*

The existence of a mixed Nash equilibrium follows from the game being a finite game, and then applying Kakutani's fixed point theorem.

4 Characterization of Pure-Strategy Nash Equilibria

In this section, we address the question of how changing the transmission parameter β affects the topology of the formed network in the equilibrium. We focus on the case where $(\mathcal{V}, E_\gamma)$ does not contain any cycle. Using Theorem 1, this network has a pure-strategy Nash equilibrium. We show that as β increases, the network structure changes from a collection of long sparse chains to dense (and possibly smaller) components. Furthermore, we introduce two threshold functions using pairwise popularities and cost parameter, and show that when transmission parameter β is lower than this threshold, the edges of equilibrium network is a subset of E_γ. However when transmission parameter β is larger than this threshold, the equilibrium network will include low valuation edges as well, and the network is segregated into smaller dense components. To facilitate the statement of the theorem, we first introduce some notations and definitions, which we will use in the rest of the paper.

Definition 5 (Clustering Coefficient and Triadic Closure) Consider a directed graph $G = (\mathcal{V}, E)$. Given the adjacency matrix $A = [a_{ij}]_{i,j\in\mathcal{V}}$ ($a_{ij} = 1$ if and only if $i \to j$), for any $i \in \mathcal{V}$, the individual clustering coefficient of i is defined as

$$C_i = \frac{\sum_{j\neq k} a_{ij}a_{jk}a_{ik}}{\sum_{j\neq k} a_{ij}a_{jk}}. \tag{4.1}$$

The clustering coefficient of i captures the fraction of friends of friends of agent i that are friend of i as well. Each non-zero term of the denominator of (4.1) is called a *triadic*, i.e., $i \to j \to k$, is a triadic associated with agent i. If $i \to j \to k$ and $i \to k$, then we say the triadic $i \to j \to k$ is *closed* and $i \to k$ is the *closing edge of the triadic.* For instance, if all the triadics are closed, then the clustering coefficient of each agent is one.

We define *minimum connection loss*, denoted by ι_M, as the minimum direct "damage" an agent incurs when establishing a low value connection, i.e.,

$$\iota_M = \gamma - \max\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i,j \in \mathcal{V}\},$$

and we define *maximum connection loss*, denoted by ι_m, as the maximum direct "damage" an agent occurs when establishing a low value connection, i.e.,

$$\iota_m = \gamma - \min\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i,j \in \mathcal{V}\}.$$

We refer to $\beta\gamma$ as the *minimum indirect gossip cost.* We also refer to $\gamma(1-(1-\beta)^\mu)$ as the *maximum indirect gossip cost*, where μ is the maximum min-cut among all pairs of nodes in $(\mathcal{V}, E_\gamma)$, defined as

$$\mu = \max_{i,j \in \mathcal{V}} \mu(i,j),$$

where $\mu(i,j)$ is the minimum number of edges whose removal will disconnect i from j in the graph $(\mathcal{V}, E_\gamma)$.

Using these definitions, we can state the phase transition result as follows:[7]

Theorem 2 (Phase Transition in Equilibrium Network) *Given valuation matrix* $\mathbf{V}$*, suppose popular-connections does not have any cycle. In the equilibrium networks, we have*

1. *If* maximum indirect gossip cost *is lower than* minimum connection loss*, i.e.,* $\gamma(1-(1-\beta)^\mu) < \iota_M$*, then the edges of equilibrium network is a subset of* E_γ*, and the individual clustering coefficient of each agent is at most the individual clustering coefficient of that node in the popular-connections graph.*

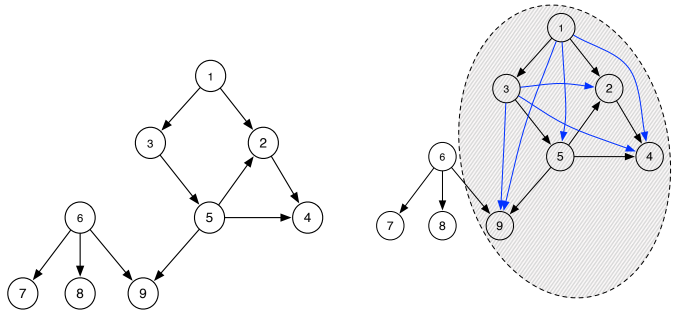
2. *If* minimum indirect gossip cost *is higher than* maximum connection loss *, i.e.,* $\beta\gamma > \iota_m$*, then in equilibrium network the individual clustering coefficient of each agent is one.*

The intuition for this theorem can be obtained by noting that the *minimum connection loss* is a lowerbound on the damage that connecting to an agent can cause (not including indirect effects), while the *maximum connection loss* is an upperbound. Therefore, when the maximum cost of indirect gossip, $\gamma(1-(1-\beta)^\mu)$, is less than the minimum connection loss, then the equilibrium network will never contain an agent who is not present in the popular-connections graph (meaning that any edge for which the indirect cost of gossip exceeds the direct benefit). Conversely, when the minimum cost of indirect gossip, $\gamma\beta$, is greater than the maximum connection loss, an agent will always connect to a friend of her friend, even if she has a low direct benefit from this connection, because the cost of indirect gossip always exceeds this value, implying that it is always better to make a direct connection and obtain the benefit of this connection rather than suffer indirect gossip. This reasoning yields the result that all triadics will be closed and thus the individual clustering coefficient of each agent will be one.

Remark 1 If $\iota_M = \iota_m$ and $\mu = 1$[8], then Theorem 2 establishes a sharp phase transition as a function of β for determination of individual clustering coefficient of all agents.

[7] Using Proposition 1 and the construction of Nash equilibrium given in Theorem 1, the equilibrium network characterize in the next theorem is also a strong Nash equilibrium.

[8] For instance, if in the graph $(\mathcal{V}, E_\gamma)$ there is only one directed path between any two nodes, then we have $\mu = 1$.



(a) $\beta < 1 - \frac{1}{\sqrt{2}}$ (b) $\beta > \frac{1}{2}$

Figure 4: Equilibrium network for the setting described in Example 3. By increasing the transmission parameter β, as Theorem 2 shows the clustering coefficient of all nodes jump to one.

The intuition for this result is instructive. As we increase β, there are two effects loosely corresponding to the forces identified before creating respectively strategic complementarity and substitutability. First, consider the decision of agent i to connect to agent j with whom he is indirectly connected to (i.e., there exists a directed path of friendship from i to j). As β increases, the probability that i's information will leak to j through indirect path increases, encouraging i to directly connect to j. The second force can be seen in the case when i considers connecting to individual j with whom she is not indirectly connected. In this case, by initiating a connection, i will expose himself to indirect leakage to j's friends. As β increases, this leakage becomes more likely discouraging connection from i to j.

When β is low, our result shows that the second force dominates and the network is sparse consisting only edges of E_γ (agents only connect to a subset of their high valuation friends). When β is large, the first force dominates and generates a network in which clustering coefficient of all nodes are one.[9]

We will demonstrate the results of Theorem 2 in the next example.

Example 3 We consider the the same setting as in Example 2, where there are 9 nodes and all the high value pairs (shown in the column at the left hand side of Figure 3) have valuation $H = \frac{9}{4}\gamma$ and all the low value pairs have valuation $L = \frac{1}{2}\gamma$. This choice of H and L guarantees that $(H-\gamma) + 2(L-\gamma) > 0$ and $(H-\gamma) + 3(L-\gamma) < 0$, which is the same assumption as in Example 2. Since $\iota_M = \iota_m = \frac{1}{2}\gamma$, and $\mu = 2$, the thresholds on β predicted by Theorem 2 becomes $\beta < 1 - \frac{1}{\sqrt{2}}$ and $\beta > \frac{1}{2}$. For $\beta < 1 - \frac{1}{\sqrt{2}}$, the equilibrium network is depicted in Figure 4a, as predicted by part (a) of Theorem 2. For any $\beta > \frac{1}{2}$,

[9]The phase transition can also be viewed as the effect of changing the cost parameter γ on the topology of network equilibrium. Because the conditions $\beta\gamma < \iota_M$ and $\beta\gamma > \iota_m$ can be viewed as conditions on γ.

the equilibrium network is depicted in Figure 4b, as predicted by part (b) of Theorem 2. For instance the clustering coefficient of nodes 1 and 3 with $\beta < 1 - \frac{1}{\sqrt{2}}$ (Figure 4a) are equal to 0, whereas the clustering coefficient of both nodes 1 and 3 with $\beta > \frac{1}{2}$ (Figure 4b) are equal to 1.

5 Triadic Closure and Homophily

In this section we show the emergence of triadic closure and high clustering coefficient in equilibrium network. Note that the results of this section hold for any valuation matrix $\mathbf{V}$, i.e., we do not impose any assumption on the popular-connections graph.

5.1 Triadic Closure in Equilibrium Network

In the next theorem, we consider triadic closure in both pure-strategy Nash equilibrium and mixed Nash equilibrium and establish sufficient conditions under which the triadics are closed in the equilibrium.

Theorem 3 (Triadic Closure in Equilibrium) *Given $\beta \in [0, 1]$, in any pure-strategy Nash equilibrium $\mathbf{x}$, when $x_{ij} = 1$ and $x_{jk} = 1$, then $x_{ik} = 1$ if*

$$v_{ik} \geq \gamma(1 - \beta)(1 + \beta \ \mathit{Gossip}(k, \mathbf{x})), \tag{5.1}$$

i.e., triadics are closed if pairwise valuation of the closing edge is higher than the marginal increase of the gossip cost. Similarly, in any mixed Nash equilibrium $\boldsymbol{\sigma}$, if for $i, j, k \in \mathcal{V}$, $\mathbb{P}_\sigma[x_{ij} = 1, x_{jk} = 1] = 1$, then $\mathbb{P}_\sigma[x_{ik} = 1] = 1$, if

$$v_{ik} \geq \gamma(1 - \beta) \left(1 + \beta \ \mathbb{E}[\mathit{Gossip}(k, \boldsymbol{\sigma})]\right).$$

Theorem 3 thus shows that, provided that condition (5.1) is satisfied, all triadics will be closed. This condition is in fact not very restrictive, because the term $\gamma(1 - \beta)(1 + \beta \ \text{Gossip}(k, X))$ will be typically small. In particular, it will tend to be small when

(a) β is close to 1, which implies that the existence of edge $i \to k$ will slightly change the cost of gossip for i, since with a high probability the information of i has already been leaked to k through the path $i \to j \to k$.

(b) Gossip (k, X) is small, implying that making a connection to k would leak a small amount of information. Hence, the benefit of connection to k overcomes the cost of gossip by forming a connection to k and connection happens if $v_{ik} > \gamma$.

(c) γ is small which implies that the cost of gossip is small, so that agent i would connect to agent k.

Also, note that when $\beta = 1$, the right-hand side of condition (5.1) will be zero, which implies that in any equilibrium triadics will be closed.

Corollary 1 *For transmission parameter $\beta = 1$, in any pure-strategy Nash equilibrium all triadics are closed, i.e., for three nodes, i, j, k, if $x_{ij} = 1$ and $x_{jk} = 1$, then $x_{ik} = 1$. Similarly, in the mixed Nash equilibrium $\boldsymbol{\sigma}$, for $i, j, k \in \mathcal{V}$, if we have $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ij} = 1, x_{jk} = 1] = 1$, then $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ik} = 1] = 1$.*

This corollary also implies that if there is any path (not necessarily a triadic) that delivers i's information to k, then i must be connected to k. In other words, in the equilibrium if there exists a path $i = i_0 \to i_1 \to \cdots \to i_l = k$, i.e., $x_{i_j i_{j+1}} = 1$, for all $j \in \{0, \ldots, l-1\}$, then we would have $x_{ik} = 1$.

Condition (5.1) implies when β is large in any equilibrium all triadics are closed. We next argue that even if β is small, the existence of large connected communities will induce triadic closure in any equilibrium. The reason is that if an agent connects to a member of a large connected community, then her information leaks to all members of that community with a probability close to one, and this implies that if she connects to any agent in a large community, she would prefer to connect to all members of that community. This ensures that all the triadics will be closed.

Remark 2 (Effect of Community Size) Suppose that the set of agents $\mathcal{V}$ is segregated into several communities, where the pairwise valuation in each community is high enough that it encourages all the members of each community to connect to each other in any equilibrium network. Let C denote the minimum size of these communities. Using Corollary 1 if $\beta = 1$, then all triadics are closed. We now argue that for any $\beta > 0$, if C is large enough, then again in any equilibrium network, all the triadics will be closed, i.e., the equilibrium network becomes a completely clustered network. This observation means that large communities will be either completely disconnected from each other or completely connected to each other. Intuitively, and anticipating the reasoning we will employ in our analysis of homophily in the next section, if an agent a in community A connects to an agent b in community B, then her information will leak to all members of community B, thus a will form a connection to all members of B. On the other hand, all members of A are connected to a and therefore, their information also leaks to all members

of B through a, resulting in two completely connected communities.
In particular, we show that if

$$C \geq \left(\frac{1}{\log(\frac{1}{1-\beta^2})}\right) \log\left(\frac{n\gamma}{\ell_0(1-\beta^2)^2}\right), \tag{5.2}$$

where $\ell_0 = \min\{v_{ij} \mid i \neq j,\ i, j \in \mathcal{V}\}$ is nonzero, then all triadics will be closed.
We first argued that if agent a in community A connects to b in B, then it must connect to all agents in community B. The reason is that for any $b' \in B$, we have

$$\mathbb{P}[a \rightsquigarrow b'] \geq \left(1 - (1-\beta^2)^{C-2}\right),$$

because there are at least $C-2$ distinct paths from a to b' each with probability β^2. With probability $\mathbb{P}[a \rightsquigarrow b']$ two terms $\text{Gossip}(a, \mathbf{x}, x_{ab'} = 1)$ and $\text{Gossip}(a, \mathbf{x}, x_{ab'} = 0)$ are equal and with probability $1 - \mathbb{P}[a \rightsquigarrow b']$, their difference is at most n, which yields to

$$\begin{aligned}
&\text{Gossip}(a, \mathbf{x}, x_{ab'} = 1) - \text{Gossip}(a, \mathbf{x}, x_{ab'} = 0) \\
&\leq \mathbb{P}[a \rightsquigarrow b'] \times 0 + \left(1 - \mathbb{P}[a \rightsquigarrow b']\right) \times n \\
&\leq n\left(1 - \left(1 - (1-\beta^2)^{C-2}\right)\right) = n(1-\beta^2)^{C-2}.
\end{aligned}$$

Let $\tilde{u}_a$ shows the utility of a after deviation and connecting to node b' in B. We have

$$\tilde{u}_a - u_a \geq v_{ab'} - \gamma n(1-\beta^2)^{C-2},$$

which is positive provided that

$$v_{ab'} \geq \gamma n(1-\beta^2)^{C-2},$$

and this holds under the condition given in (5.2).
The second claim is that since a in A is connected to all nodes in B, all other agents in A will also connect to all nodes in B. This follows from the same argument. In particular, since there are many paths from a' in A to a that will leak a''s information to all nodes in B with probability close to one (since C is large), a' would have a profitable deviation with connecting to all nodes in B herself. These two claims show that the clustering coefficient has to be one.

5.2 Homophily

As noted above, homophily refers to a situation in which agents are more likely to be friends with or have links with others in their community than those outside the community. In this subsection, we show that when there is a slight difference in terms of direct benefits from connecting within the community, this will lead to a significant pattern of homophily because of privacy concerns.

To establish these results, we will focus on the strong Nash equilibrium of the network formation game because, as the next example illustrates, there will sometimes also exist other, unintuitive pure-strategy Nash equilibria, but these are never strong Nash equilibria.

Example 4 (Nash equilibrium versus strong Nash equilibrium) Suppose that we have two groups of agents each of them of size $n > 1$, denoted by A and B. Let the valuation within each group to be H and valuation across groups to be L (symmetric valuations), except that valuation between a_i and b_i are H for $i = 1, \ldots, n$. Also, let $0 < L < \gamma < H$, $\beta = 1$, and $2H + L - 3\gamma < 0$. One pure-strategy Nash equilibrium, for this network is shown in Figure 5 where a_i is connected to b_i for all $i = 1, \ldots, n$ and there is no other connection between agents. However, it is not a strong Nash equilibrium, because if all agents $a_1, \ldots, a_n$ deviate together, disconnect from B and connect to each other, then the utility of each of them would be improved to $n(H - \gamma)$, instead of $H - \gamma$ in the Nash equilibrium shown in Figure 5. The strong Nash equilibrium network for this setting is two segregated completely connected communities as depicted in figure 5.
Intuitively, in strong Nash equilibrium, privacy concerns force the society to form two separate communities. The reason is that if an agent such as a is part of a community A and the rest of community A is connected to another community B, then a would benefit from connection to B as well, since her information is already leaking to agents in B through her connections with people of her own community. However, if nodes in A decide altogether to disconnect from B, then this would be beneficial for all of them. This example thus illustrates why the notion of strong Nash equilibrium plays an important, albeit intuitive, role in the emergence of clustered networks and homophily.

Example 4 motivates our focus on strong Nash equilibria in the rest of this subsection. We will then see that strong Nash equilibria will feature a strong form of homophily.

Let us now focus on the two-community society (with the two groups denoted by A and B) with the following probablilistic popularity pattern:[10]

$$v_{ij} = \begin{cases} L & \text{w.p. } p \quad \text{if } i,j \in \text{ same group} \\ H & \text{w.p. } 1-p \quad \text{if } i,j \in \text{ same group} \\ L & \text{w.p. } 1-p \quad \text{if } i,j \in \text{ different groups} \\ H & \text{w.p. } p \quad \text{if } i,j \in \text{ different groups}, \end{cases}$$

[10]Though we describe this popularity pattern probabilistically, at the time of forming connections, agents have complete information about popularities.

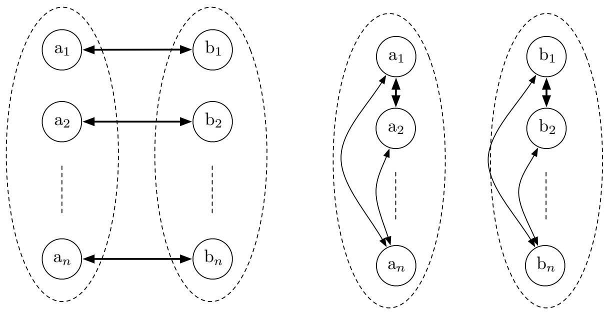


Figure 5: Nash equilibrium vs. strong Nash equilibrium

where L corresponds to low popularity, and $H > L$ to high popularity, and $p \in [0,1]$. These preferences thus indicate that there is some "homophily" in preferences, but this is quite weak because H could be arbitrarily close to L. More specifically, within each community, agents have on average pn low popularities and across the communities agents have on average pn high popularities. However, we show that that even in this setting the equilibrium network will be highly clustered and will feature homophily. We should emphasize that, though we are considering a probabilistic setting in terms of the valuation matrix, the game is still one of complete information, i.e., the players know the $\mathbf{V}$ matrix when making their choices.

Theorem 4 *Let $\beta \in (0,1]$ and consider the probability distribution of v_{ij}'s described above. For any $\eta > 0$ there exist n_0 such that for any $n \geq n_0$ with probability at least $1-\eta$:*

(a) If $H, L > \gamma$, then a complete network is the only strong Nash equilibrium.

(b) If $L < \gamma < H$ and $a = \frac{\gamma - L}{H-\gamma} \geq 0$, then for $p < \min\{\frac{1}{(a+1)^2+a}, \frac{1}{2a+3}, \frac{1}{\frac{H-\gamma}{\gamma\beta}+a+3}\}$, any strong Nash equilibrium contains two segregated groups, i.e., there is no cross community connection. Furthermore, there exist completely connected sub-groups $S_A \subseteq A$ and $S_B \subseteq B$ of size at least $n(a+2)(\frac{1}{a+2} - p)$, where all agents of $A \setminus S_A$ are connected to all agents of S_A and all agents of $B \setminus S_B$ are connected to all agents of S_B.

Thus a very strong form of homophily, with two segregated communities, emerges as the unique strong Nash equilibrium. The intuition is as follows: because of the slight preference for within-community links, there will be more within-community connections, but this in turn implies that an agent will have further incentives to form within-community links because her information is already likely to have leaked to her potential friends within the community. In contrast, she will refrain from links to the other community, because even a single across-community link will imply the leakage of her information to many other connected people within this other community.

Note also that the parameter p needs to be smaller than a certain threshold for this result to hold. Clearly, p needs to be less than $1/2$, but the theorem specifies a lower threshold, which helps ensure that there are sufficiently many within-community links than cross-community links in any (strong) Nash equilibrium.

6 Conclusion

The technological developments of the last decade and a half have increased the sharing of information over various social media. With this trend set to continue, concerns over privacy have also mounted, and are expected to become a growing constraint on the functioning of many online platforms. Despite the centrality of issues of privacy both in online platforms and in various real-world and virtual social networks, there is relatively little game-theoretic analysis of privacy and efforts by agents to protect their privacy.

In this paper, we took the first step in this direction by modeling how privacy concerns affect individual choices in the context of a network formation game (where links can be interpreted as friendships in a social network, connections over a social media platform or trading activity in online platform). In the model, each individual decides which other agents to "befriend", i.e., form links with. Such links bring direct (heterogeneous) benefits from friendship and also lead to the sharing of information. But such information can travel over other linkages (e.g., shared by the party acquiring the information with others), defining a percolation process over the equilibrium network. Privacy concerns are modeled as a disutility that individual suffers as a result of her private information being acquired by others, and imply that the individual has to take into account who the friends of her new friend (and who the friends of friends of her new friend etc.) are.

After showing that pure-strategy Nash equilibria may fail to exist, we provided sufficient (and necessary) conditions for the existence of pure-strategy equilibria, and characterized their structure. Information flows over the social networks create both strategic complementarities and substitutabilities, which in turn lead to a phase transition result whereby small changes in the transmission of information over the network can fundamentally change the nature (and clustering) of the equilibrium network.

Our main results concern the analysis of triadic closure and homophily. We show that clustering of links and thus triadic closure emerges naturally because if player a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, thus making it less costly for a to befriend c. Homophily also emerges as part of the equilibrium network formation (provided that we focus on strong Nash to avoid other potential equilibria with the flavor of coordination failure). This is because even an infinitesimal advantage in terms of direct benefits of friendship within a group makes linkages within that group more likely, in turn making information travel within that group, reducing the cost of making further within-group links due to loss of privacy, and thus increasing the likelihood of further within-group links.

References

Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

Chris Baerveldt, Marijtje AJ Van Duijn, Lotte Vermeij, and Dianne A Van Hemert. Ethnic boundaries and personal choice. assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils networks. *Social Networks*, 26(1):55–74, 2004.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

Hubert M Blalock. *Race and ethnic relations*. Prentice-Hall Englewood Cliffs, NJ, 1982.

Peter Michael Blau. *Inequality and heterogeneity: A primitive theory of social structure*, volume 7. Free Press New York, 1977.

Larry Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Éva Tardos. Network formation in the presence of contagious risk. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 1–10. ACM, 2011.

Giacomo Calzolari and Alessandro Pavan. Optimal design of privacy policies1. 2001.

Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002a.

Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002b.

Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88(12):128701, 2002.

Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 2. Springer, 1998.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2013.

Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H Papadimitriou, and Scott Shenker. On a network creation game. In *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pages 347–351. ACM, 2003.

Giorgio Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2): 026107, 2007.

Eric Fong and Wsevolod W Isajiw. Determinants of friendship choices in multiethnic society. In *Sociological Forum*, volume 15, pages 249–271. Springer, 2000.

James H Fowler, Christopher T Dawes, and Nicholas A Christakis. Model of genetic variation in human social networks. *Proceedings of the National Academy of Sciences*, 106(6):1720–1724, 2009.

Andrea Galeotti, Sanjeev Goyal, and Jurjen Kamphorst. Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372, 2006.

Jack Hirshleifer. Privacy: Its origin, function, and future. *The Journal of Legal Studies*, pages 649–664, 1980.

Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Physical review E*, 65(2):026107, 2002.

Kai-Lung Hui and Ivan PL Png. The economics of privacy. *Handbooks in Information System and Economics*, 1:471–493, 2006.

Matthew O Jackson. A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, pages 11–49, 2005.

Matthew O Jackson and Brian W Rogers. Meeting strangers and friends of friends: How random are social networks? *The American economic review*, pages 890–915, 2007.

Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of economic theory*, 71(1):44–74, 1996.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

Jon Kleinberg, Christos H Papadimitriou, and Prabhakar Raghavan. On the value of private information. In *Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge*, pages 249–257. Morgan Kaufmann Publishers Inc., 2001.

Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.

Kenneth C Laudon. Markets and privacy. *Communications of the ACM*, 39(9):92–104, 1996.

Paul F Lazarsfeld, Robert K Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

Charles E Leiserson, Ronald L Rivest, Clifford Stein, and Thomas H Cormen. Introduction to algorithms, 2001.

Yingbin Liang, H Vincent Poor, et al. Information theoretic security. *Foundations and Trends in Communications and Information Theory*, 5(4–5):355–580, 2009.

Trina J Magi. Fourteen reasons privacy matters: A multidisciplinary review of scholarly literature1. *The Library*, 81(2), 2011.

Peter V Marsden. Core discussion networks of americans. *American sociological review*, pages 122–131, 1987.

Peter V Marsden. Homogeneity in confiding relations. *Social networks*, 10(1):57–76, 1988.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

AD Medus and Claudio Oscar Dorso. Triadic closure mechanism in face-to-face and online social relationship networks. *arXiv preprint arXiv:1312.3496*, 2013.

James Moody. Race, school integration, and friendship segregation in america1. *American Journal of Sociology*, 107(3):679–716, 2001.

Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2): 167–256, 2003.

Mark EJ Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5200–5205, 2004.

Eleonora Patacchini and Yves Zenou. Racial identity and education. Technical report, IZA Discussion Papers, 2006.

Anatol Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, 15(4):523–533, 1953.

Pamela Samuelson. Privacy as intellectual property? *Stanford Law Review*, pages 1125–1173, 2000.

Brian Skyrms and Robin Pemantle. A dynamic model of social network formation. In *Adaptive Networks*, pages 231–251. Springer, 2009.

George J Stigler. An introduction to privacy in economics and politics. *The Journal of Legal Studies*, pages 623–644, 1980.

Eva Tardos and Tom Wexler. Network formation games and the potential function method. *Algorithmic Game Theory*, pages 487–516, 2007.

Curtis R Taylor. Consumer privacy and the market for customer information. *RAND Journal of Economics*, pages 631–650, 2004.

Hal R Varian. Economic aspects of personal privacy. *Privacy and Self-regulation in the Information Age*, 1997.

Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):056104, 2003.

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

Alan F Westin. Privacy and freedom. 1967.

A Appendix

A.1 Proof of Theorem 1

We first show a lemma that will be used in the proof.

Lemma 1 *Let $G = (\mathcal{V}, E)$ be a directed graph. For any $a, b \in \mathcal{V}$, the gossip probability $\mathbb{P}[a \rightsquigarrow b]$ increases (or does not change) if we add any edge $c \to d$.*

Proof The proof is straightforward. Let $\tilde{P}_a(G)$ be the probability of the realized graph being G, after adding edge $c \to d$, given the source of information is node a. We have that

$$\begin{aligned}
&\mathbb{P}[a \rightsquigarrow b] = \sum_{G\,:\,(c,d)\notin E_{\mathbf{x}}} P_a(G)\mathbf{1}\{a \rightsquigarrow b \text{ in the graph } G\} \\
&= \sum_{G\,:\,(c,d)\in E_{\mathbf{x}}} \mathbb{P}[x_{cd} = 0] P_a(G)\mathbf{1}\{a \rightsquigarrow b \text{ in the graph } G\} \\
&\leq \sum_{G\,:\,(c,d)\in E_{\mathbf{x}}} \left(\mathbb{P}[x_{cd} = 0] + \mathbb{P}[x_{cd} = 1]\right) P_a(G)\mathbf{1}\{a \rightsquigarrow b \text{ in the graph } G\} \\
&= \sum_{G\,:\,(c,d)\in E_{\mathbf{x}}} \tilde{P}_a(G)\mathbf{1}\{a \rightsquigarrow b \text{ in the graph } G\}.
\end{aligned}$$

This completes the proof of lemma. ■

We now proceed with the proof of Theorem.

Proof of part 1: We first prove the necessary condition by showing that if there exists a directed cycle of length at least three in popular-connections graph, then there exists a set of popularities $\mathbf{V}$ and β small enough for which pure-strategy Nash equilibrium does not exist. Given popular-connections graph, consider the smallest cycle in this network, denoted as $C = (\mathcal{V}_c, E_c)$, where $\mathcal{V}_c = \{v_1^c, v_2^c, \dots, v_k^c\}$. Without loss of generality suppose $\mathcal{V}_c = \{1, \dots, k\}$. We now assign v_{ij}'s that are consistent with $(\mathcal{V}, E_\gamma)$.

1. For all edges that do not belong to E_γ let $v_{ij} = 0$. Since we have $\gamma > 0$, this guarantees that in any Nash equilibria, there exist no edge among nodes i and j where $v_{ij} = 0$.

2. For $(i, j) \in E_\gamma$, where either $i \notin \mathcal{V}_c$ or $j \notin \mathcal{V}_c$, let $v_{ij} = \infty$. This guarantees that in any Nash equilibria $x_{ij} = 1$.

Next, we show that there exist an assignment of v_{ij}'s for $i, j \in \mathcal{V}_c$ for which no pure-strategy Nash equilibrium exist.

First note that since $(\mathcal{V}_c, E_c)$ is the smallest cycle in $(\mathcal{V}, E_\gamma)$, for all $i, j \in \mathcal{V}_c$ such that

$i \neq j + 1 \pmod{k}$, $(i, j) \notin E_\gamma$ (otherwise we can find a cycle smaller than C which contradicts the fact that C is the smallest cycle), therefore we have assigned $v_{ij} = 0$ to those pairs, implying there is no connection among $i, j \in \mathcal{V}_c$ such that $i \neq j + 1 \pmod{k}$. Let $v_{i,i+1} > \gamma$ for $i = 1, \ldots, k$ (all i's are modulus k). We now specify these popularities. We consider two cases, depending on whether k is odd or even.

(i) k is an odd number: let $\mathbf{x}$ show the connections of agent without considering connections among nodes in $\mathcal{V}_c$. We claim that for $\beta > 0$ small enough, for $i = 1, \ldots, k$, we have

$$\begin{aligned}&\min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)})\\ &> \max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)}),\end{aligned}$$

where $\mathbf{x}^c_{-(i,i+1)} = (x_{j,(j+1)} \ : \ j \neq i, i + 1)$. Next, we prove this claim.

Proof of Claim: Using Lemma 1, the claim is equivalent to

$$\begin{aligned}&\text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)} = 0)\\ &> \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)} = 1).\end{aligned}$$

For β sufficiently small, we have

$$\text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)} = 0) = \sum_{j \in \mathcal{V}} x_{ij} + \beta \sum_{j,k \in \mathcal{V}} x_{ij} x_{jk} + o(\beta^2),$$

and

$$\text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)} = 1) = \sum_{j \in \mathcal{V}} x_{ij} + \beta \sum_{j,k \in \mathcal{V}} x_{ij} x_{jk} + o(\beta^2).$$

Comparing the previous two equations, the first terms are equal and the second term of the first equation is larger that the second term of the second equation as it contains $i \to i + 1 \to i + 2$. Thus, for sufficiently small β, we obtain

$$\begin{aligned}&\text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)} = 0)\\ &> \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)} = 1).\end{aligned}$$

This completes the proof of claim. ■

Back to the proof of Theorem. For any $i = 1, \ldots, k$, we let

$$\begin{aligned}v_{i,i+1} \in (&\gamma \max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)}),\\ &\gamma \min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)})).\end{aligned}$$

We now completed matrix $\mathbf{V}$ which is compatible with popular-connections graph. We next show that there exists no pure-strategy Nash equilibrium for valuation matrix $\mathbf{V}$. First, note that if there exists an equilibrium, we can find i, such that $i \to i+1$ (because a network with no edge on the cycle is not an equilibrium as one of the agents such as i can deviate and connect to $j = i+1$ (mode k)). Without loss of generality, suppose $k \to 1$. This would imply that $k-1 \not\to k$, since agent $k-1$ faces a loss equal to

$$\gamma \text{Gossip}(k-1, \mathbf{x}, x_{k,1} = 1, x_{k-1,k} = 1, \mathbf{x}^{*c}),$$

compared to the benefit of $v_{k-1,k}$. In other words, if $k-1$ connects to k, her utility would be smaller than

$$\gamma(\min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(k, X x_{k,1} = 1, x_{k-1,k} = 1, \mathbf{x}^c_{-(i,i+1)}) \\ - \text{Gossip}(k, X, x_{k,1} = 1, x_{k-1,k} = 1, \mathbf{x}^{*c})),$$

which is non-positive.
This in turn shows the following set of decisions $k-2 \to k-1, k-3 \not\to k-2, \ldots, 1 \to 2, k \not\to 1$, which is a contradiction (we started with the assumption $k \to 1$), showing that no pure-strategy Nash equilibrium exists for the assigned pairwise popularities and β small enough.

(ii) If k is even, similar to the proof of claim, we can show

$$\max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)}) \\ < \min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)}),$$

and

$$\max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(k, \mathbf{x}, x_{12} = 1, x_{k,1} = 1, x_{2,3} = 0, \mathbf{x}^c_{-(i,i+1)}) > \\ \min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{k,1} = 1, x_{2,3} = 1, \mathbf{x}^c_{-(i,i+1)}).$$

Now let $v_{1,2} = \infty$, and for $i = 2, \ldots, k-1$ let

$$v_{i,i+1} \in (\gamma \max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c_{-(i,i+1)}), \\ \gamma \min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c_{-(i,i+1)})),$$

along with

$$v_{k,1} \in (\gamma \max_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(k, \mathbf{x}, x_{12} = 1, x_{k,1} = 1, x_{2,3} = 0, \mathbf{x}^c_{-(i,i+1)}),$$
$$\gamma \min_{\mathbf{x}^c_{-(i,i+1)}} \text{Gossip}(i, \mathbf{x}, x_{12} = 1, x_{k,1} = 1, x_{2,3} = 1, \mathbf{x}^c_{-(i,i+1)})).$$

We now consider two scenarios:

- Either $2 \to 3$: This would imply that $k \not\to 1$, which in turn shows that $k - 1 \to k, \ldots, 3 \to 4, 2 \not\to 3$, which is to a contradiction,
- Or $2 \not\to 3$: This would imply that $k \to 1$, which in turn shows that $k - 1 \not\to k, \ldots, 3 \not\to 4, 2 \to 3$, which is again a contradiction, completing the proof.

Therefore, we showed that if there exists a cycle in popular-connections, we can construct $\mathbf{V}$ compatible with popular-connections, for which there exists no pure-strategy Nash equilibria (for β small enough).

Proof of part 2: We next show that the absence of cycles in popular-connections is sufficient for the existence of pure-strategy Nash equilibrium. We will construct an update rule that converges to an equilibrium in finitely many steps. The construction is as follows. Since the graph popular-connections has no cycle, there exist some nodes with no outgoing edges (otherwise, we would have a cycle). Let R_1 denote the set of nodes with no outgoing edges. Also let $H_1 = (\mathcal{V}, E_\gamma)$. In each step of the construction, we update the set R and the graph H. In the first step of the construction, we let all nodes to play their optimal decision regarding the nodes in R_1. We now update H_1 by adding the newly created edges to obtain H_2. We then update R_1 to be the set of nodes with no outgoing edges plus the nodes with outgoing edges only in the set R_1, and denote the updated set by R_2. We now proceed to the next step of construction. In the second step, we let all nodes to play their optimal decision regarding the nodes in R_2. We then update H_2 by adding the newly created edges. Finally, we update R_2 to be the set of nodes with outgoing edges only in the set R_2 as well as the nodes with no outgoing edges. We continue the steps until R becomes the entire set of nodes $\mathcal{V}$. We claim that the resulting graph is an equilibrium network.

We first show that a node such as a that is about to be added to R at step r, will only (potentially) connect to nodes in the set R_r. We will show this by induction on r. For $i = 1$ it is evident. Suppose it holds for all $i \leq r - 1$. We will show that it holds for $i = r$ as well. Since node a does not have any high value friend in $\mathcal{V} \setminus R_r$ and non of the nodes in R_r is connected to nodes in $\mathcal{V} \setminus R_r$ (by induction assertion), node a will only make connections to nodes in R_r. Therefore, it suffices to show that a will not deviate from the optimal decision she makes at step r. This is evident as a is playing his best response regarding connection to nodes in R_r and all the nodes in R_r will not change their connections in the subsequent steps.

Next, we show that best response dynamics converges to equilibrium. In the first round of best response dynamic all the nodes will play their optimal decision regarding the nodes in the set R_1, same as the first step of the construction, and do not change their decisions in the rest of the best response dynamics. Since, nodes do not change their decisions regarding connection to nodes in R_1, in the second round, all of the nodes will play their best response regarding connection to nodes in R_2 and do not change their decisions in the next rounds. By repeating this argument, after finite number of rounds the created edges would be the same as the ones that the construction steps described before, would create. This shows that the resulting network after finitely many best response dynamic rounds is a pure-strategy Nash equilibrium.

A.2 Proof of Proposition 1

We use the same notation as in the proof of Theorem 1. Consider the constructed Nash equilibrium in the proof of part 2 of Theorem 1. We will show that it is a strong Nash equilibrium. Consider a set of agents $S \subseteq \mathcal{V}$. We will show that the group S of agents does not have a profitable deviation. Note that no matter what the decisions of other nodes are, the nodes in $S \cap R_1$ will not change their decisions. Now that these nodes do not change their decisions, all nodes in $S \cap R_2$ will not change their decisions. By repeating this argument, none of the nodes in $\cup_r (S \cap R_r) = S$ (as $\cup_r R_r = \mathcal{V}$) will change their decisions, showing that the resulting equilibrium in Theorem 1 is a strong Nash equilibrium.

A.3 Proof of Theorem 2

Let

$$\ell_M = \max\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i,j \in \mathcal{V}\},$$

and

$$\ell_m = \min\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i,j \in \mathcal{V}\}.$$

(a) Consider the construction process described in Theorem 1. We show that no edge from $\mathcal{V} \times \mathcal{V} \setminus E_\gamma$ will be included in equilibrium. We show this by following the procedure described in Theorem 1. We use the same notation as the one used in the proof of Theorem 1 given in subsection A.1. By induction on r, we will show that at step r no node will connect to a node in $\mathcal{V} \times \mathcal{V} \setminus E_\gamma$. Suppose this holds for all steps up to step i. We will show that it holds for step i as well. Suppose that node a in step i wants to connect to node $b \in R_i$ that is not a neighbor of a in $(\mathcal{V}, E_\gamma)$. Also, suppose b is the last node aded to R with this property. We have $v_{ab} < \gamma$. If a connects to b, then we have

$$\text{Gossip }(a, \mathbf{x}, x_{ab} = 1) - \text{Gossip }(a, X, x_{ab} = 0) \geq (1 - \beta)^\mu,$$

where μ is the maximum min-cut among all pairs of nodes in $(\mathcal{V}, E_\gamma)$. The reason is based on the following three facts:

1) $\mathbb{P}[a \rightsquigarrow b | \mathbf{x}, x_{ab} = 1] = 1$ as there is a direct link from a to b.
2) $\mathbb{P}[a \rightsquigarrow b | \mathbf{x}, x_{ab} = 0] \leq 1 - (1 - \beta)^\mu$ by the induction assertion that there exist no edges beside the ones in E_γ for all steps $r \leq i$ and b is the last node added to set R.
3) $\mathbb{P}[a \rightsquigarrow c | \mathbf{x}, x_{ab} = 1] \geq \mathbb{P}[a \rightsquigarrow c | \mathbf{x}, x_{ab} = 0]$, for any c as more connections make the gossip more probable (Lemma 1).

Therefore, the utility of i by this connection would change by

$$\begin{aligned}
&u_a(\mathbf{x}, x_{ab} = 1) - u_a(\mathbf{x}, x_{ab} = 0) \\
&\leq (v_{ab}) - \gamma\left(\text{Gossip }(a, \mathbf{x}, x_{ab} = 1) - \text{Gossip }(a, \mathbf{x}, x_{ab} = 0)\right) \\
&\leq (v_{ab}) - \gamma(1 - \beta)^\mu \leq \ell_M - \gamma(1 - \beta)^\mu < 0.
\end{aligned}$$

This shows that the resulting Nash equilibrium only contains a subset of the edges in E_γ.

(b) We show that by following the construction given in the proof of Theorem 1, all triadics are closed. We show by induction that by following the procedure given in the proof of Theorem 1 at step r, all nodes must have closed all the possible triadics, for any r. Suppose that this statement holds for $r = i$, we will show that it holds for $r = i + 1$. Suppose at step r, node a is connected to $b \in R_r$, b is connected to $c \in R_r$, and $a \not\rightarrow c$. Also, suppose that c is the first node added to the set R with this property. Since $a \not\rightarrow c$ and $a \rightarrow b \rightarrow c$, then we have that $\mathbb{P}[a \rightsquigarrow c] \geq \beta$. Next we show that a can make a profitable deviation and connect to c. If a connects to c, then $\mathbb{P}[a \rightsquigarrow c] = 1$ and a receives a direct benefit equal to v_{ac} from connecting to c. Moreover, since c is the first node with this property, we have

$\mathbb{P}[a \rightsquigarrow d | x_{ac} = 1] = \mathbb{P}[a \rightsquigarrow d | x_{ac} = 0]$, for all d. Therefore, if a connects to c, his gain would be positive as

$$u_a(\mathbf{x}, x_{ac} = 1) - u_a(\mathbf{x}, x_{ac} = 0) =$$
$$v_{ac} - \gamma\left(\text{Gossip}(a, \mathbf{x}, x_{ac} = 1) - \text{Gossip}(a, \mathbf{x}, x_{ac} = 0)\right) \geq \ell_m - \gamma(1-\beta) > 0.$$

This shows that a would connect to c as well. Therefore, all triadics are closed in the equilibrium network.

A.4 Proof of Theorem 3

We first show a lemma that we will use in the proof.

Lemma 2 *Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a pure-strategy Nash equilibrium for a given set of v_{ij}'s and a given β. For any i and j, we have $x_{ij} = 1$ if and only if*

$$v_{ij} - \gamma\left(\textit{Gossip}\ (i, \mathbf{x}, x_{ij} = 1) - \textit{Gossip}\ (i, \mathbf{x}, x_{ij} = 0)\right) \geq 0.$$

Proof We have $x_{ij} = 1$ if and only if

$$u_i(\mathbf{x}, x_{ij} = 1) - u_i(\mathbf{x}, x_{ij} = 0) > 0.$$

Therefore, $x_{ij} = 1$ if and only if

$$v_{ij} - \gamma\left(\text{Gossip}(i, \mathbf{x}, x_{ij} = 1) - \text{Gossip}(i, \mathbf{x}, x_{ij} = 0)\right) > 0.$$

This completes the proof of lemma. ■

We now proceed with the proof of Theorem.

(a) Using Lemma 2, $x_{ik} = 1$ if and only if

$$v_{ik} \geq \gamma\left(\text{Gossip}(i, \mathbf{x}, x_{ik} = 1, x_{ij} = 1, x_{jk} = 1) - \text{Gossip}(i, \mathbf{x}, x_{ik} = 0, x_{ij} = 1, x_{jk} = 1)\right).$$

We also have that

$$\begin{aligned}
&\text{Gossip}\ (i, \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1) - \text{Gossip}\ (i, X, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0)\\
&= \sum_{l \in \mathcal{V}} \mathbb{P}[i \rightsquigarrow l | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1] - \mathbb{P}[i \rightsquigarrow l | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0]\\
&= (\mathbb{P}[i \rightsquigarrow k | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1] - \mathbb{P}[i \rightsquigarrow k | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0])\\
&+ \sum_{\substack{l \in \mathcal{V}\\ l \neq k}} \mathbb{P}[i \rightsquigarrow l | X, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1] - \mathbb{P}[i \rightsquigarrow l | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0]\\
&\overset{\text{(i)}}{\leq} (1-\beta) + \sum_{\substack{l \in \mathcal{V}\\ l \neq k}} \mathbb{P}[i \rightsquigarrow l | X, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1] - \mathbb{P}[i \rightsquigarrow l | \mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0]\\
&\overset{\text{(ii)}}{\leq} (1-\beta) + (1-\beta)\beta\text{Gossip}(k, \mathbf{x}),
\end{aligned}$$

where we used $\mathbb{P}[i \rightsquigarrow k|\mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1] = 1$ and $\mathbb{P}[i \rightsquigarrow k|\mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0] \geq \beta$ in (i) and the following inequality in (ii). With probability β, $j \to k$, which makes the two terms $\mathbb{P}[i \rightsquigarrow l|\mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 1]$ and $\mathbb{P}[i \rightsquigarrow l|\mathbf{x}, x_{ij} = 1, x_{jk} = 1, x_{ik} = 0]$ equal. With probability $1-\beta$, $j \not\to k$ the difference between the two terms is bounded by the gossip of k with the extra factor of β to account for the first hop going out from k.

Therefore, if $v_{ik} \geq \gamma(1-\beta)(1+\beta\text{Gossip}(k, \mathbf{x}))$, then we have $x_{ik} = 1$.

(b) Let $\mathbf{x}_i^* \in \text{support}(\sigma_i)$. We have

$$\mathbf{x}_i^* \in \text{argmax}_{\mathbf{x}_i \in \{0,1\}^n} u_i(\mathbf{x}_i, \boldsymbol{\sigma}_{-i}),$$

where

$$u_i(\mathbf{x}_i, \boldsymbol{\sigma}_{-i}) = \mathbb{E}_{\boldsymbol{\sigma}}[u_i(\mathbf{x}_i, \mathbf{x}_{-i})] = \sum_{j \neq i} v_{ij} x_{ij} - \gamma \mathbb{E}_{\boldsymbol{\sigma}}[\text{Gossip}(i, \mathbf{x})].$$

Using Lemma 2, we have $x_{ik}^* = 1$ if and only if

$$v_{ik} - \gamma \mathbb{E}_{\boldsymbol{\sigma}}[\text{Gossip}(i, \mathbf{x}, x_{ik}^* = 1) - \text{Gossip}(i, \mathbf{x}, x_{ik}^* = 0)] \geq 0.$$

Since $\mathbb{P}_{\sigma}[x_{ij} = 1, x_{jk} = 1] = 1$, with probability one, using the same argument as in part (a), we obtain

$$\mathbb{E}_{\boldsymbol{\sigma}}[\text{Gossip}(i, \mathbf{x}, x_{ik}^* = 1) - \text{Gossip}(i, \mathbf{x}, x_{ik}^* = 0)] \leq (1-\beta) + (1-\beta)\beta \mathbb{E}_{\boldsymbol{\sigma}}[\text{Gossip}(k, \boldsymbol{\sigma})].$$

Therefore, if

$$v_{ik} \geq (1-\beta) + (1-\beta)\beta \mathbb{E}_{\boldsymbol{\sigma}}[\text{Gossip}(k, \boldsymbol{\sigma})],$$

then for any $\mathbf{x}_i^* \in \text{support}(\sigma_i)$, we have $x_{ik}^* = 1$, which in turn shows $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ik} = 1] = 1$.

A.5 Proof of Theorem 4

(a) Since both L and H are greater than γ, $x_{ij} = 1$ for all $i, j \in \mathcal{V}$ is the only Nash equilibrium.

(b) We use the following Chernoff-Hoeffding bound in this proof (see e.g. Dembo and Zeitouni (1998)).

Lemma 3 *Let $Z_1, \ldots, Z_n$ be independent Bernoulli ($Z_i \in \{0,1\}$) random variables with $\mathbb{P}[Z_i = 1] = p_i$. Then for any $0 < \delta < 1$, we have that*

$$\mathbb{P}[\sum_{i=1}^{n} Z_i \leq (1-\delta)\sum_{i=1}^{n} p_i] \leq \exp\left(-\frac{\delta^2}{3}\sum_{i=1}^{n} p_i\right),$$
$$\mathbb{P}[\sum_{i=1}^{n} Z_i \geq (1+\delta)\sum_{i=1}^{n} p_i] \leq \exp\left(-\frac{\delta^2}{3}\sum_{i=1}^{n} p_i\right).$$

We also use the following lemma in our proof.

Lemma 4 *If the outdegree of any node of a directed graph G is greater than or equal to d, then the graph has a strongly connected component of size d.*

Proof Consider the decomposition of the directed graph into strongly connected components[11]. Form the component directed graph, also known as the condensation of the graph, by specifying one vertex v_C for each strongly connected component of G and edges those pairs (v_C, v_D) such that there is an edge in G from a vertex of C to a vertex of D. Note that if C and C' are two strongly connected components of a directed graph G and if there is a path from $u \in C$ to $u' \in C'$, then there cannot be a path from any $v' \in C'$ to any $v \in C$. Therefore, the component directed graph is a directed acyclic graph.

Since the component directed graph is an acyclic directed graph, it has a node that has zero outdegree. Otherwise, if each node has both incoming and outgoing edges, then the graph would have a directed cycle. Consider the corresponding connected component of this node. Since, this connected component is not connected to any node outside, and the out-degree of each node is greater than or equal to d, the size of this connected component must be greater than or equal to d, which completes the proof. ■

Given these lemmas, we now prove the theorem. For any $a \in A$, let $r_a^H(A)$ denote the ratio of H's among all $v_{aa'}$ for $a' \in A$. Similarly, let $r_a^L(B)$ denote the ratio of Ls among all v_{ab}'s for $b \in B$. Similarly, define $r_b^H(B)$ and $r_b^L(A)$. Let $\delta > 0$ be a number that will specify later.

Define random variable $Z_a(A) = \mathbf{1}\{r_a^H(A) \geq (1-p-\delta)\}$ associated with node a and note that the random variable associated with a is independent from the random

[11] A strongly connected component of a directed graph is a subgraph that is strongly connected, and is maximal with respect to strong connectivity

variable associated with any other node. Similarly, we define $Z_b(B) = \mathbf{1}\{r_b^H(B) \geq (1-p-\delta)\}$. Also define $Z_a(B) = \mathbf{1}\{r_a^L(B) \geq (1-p-\delta)\}$, and $Z_b(A) = \mathbf{1}\{r_b^L(A) \geq (1-p-\delta)\}$. Using Lemma 3, for any a, we have that

$$\begin{aligned}\mathbb{P}\left[Z_a(A)=0\right] &= \mathbb{P}\left[r_a^H(A) \leq (1-p-\delta)\right] \\ &= \mathbb{P}\left[\sum_{a' \in A} \mathbf{1}\{v_{aa'} = H\} \leq (n-1)(1-p-\delta)\right] \\ &\leq \exp\left(-(n-1)\frac{\delta^2}{3(1-p)}\right). \qquad (A.1)\end{aligned}$$

Similarly, for any a, we have that

$$\begin{aligned}\mathbb{P}\left[Z_a(B)=0\right] &= \mathbb{P}\left[r_a^L(B) \leq (1-p-\delta)\right] \\ &= \mathbb{P}\left[\sum_{b \in B} \mathbf{1}\{v_{ab} = L\} \leq n(1-p-\delta)\right] \\ &\leq \exp\left(-n\frac{\delta^2}{3(1-p)}\right). \qquad (A.2)\end{aligned}$$

Define the event

$$E = \cap_{a \in A}\{Z_a(A)=1\} \cap_{b \in B} \{Z_b(B)=1\} \cap_{a \in A} \{Z_a(B)=1\} \cap_{b \in B} \{Z_b(B)=1\}.$$

Using union bound along along with eq. (A.1) and (A.2), we obtain

$$\begin{aligned}\mathbb{P}[E] &= \mathbb{P}\left[\cap_a\{Z_a(A)=1\} \cap_b \{Z_b(B)=1\} \cap_a \{Z_a(B)=1\} \cap_b \{Z_b(B)=1\}\right] \\ &= 1 - \mathbb{P}\left[\cup_a\{Z_a(A)=0\} \cup_b \{Z_b(B)=0\} \cup_a \{Z_a(B)=0\} \cup_b \{Z_b(B)=0\}\right] \\ &\geq 1 - \left((n-1)\mathbb{P}\left[Z_a(A)=0\right] + (n-1)\mathbb{P}\left[Z_b(B)=0\right] + n\mathbb{P}\left[Z_a(B)=0\right] + n\mathbb{P}\left[Z_b(A)=0\right]\right) \\ &\geq 1 - 4n\exp\left(-(n-1)\frac{\delta^2}{3(1-p)}\right). \qquad (A.3)\end{aligned}$$

Therefore, for any η there exist $n(\eta)$ such that for any $n \geq n(\eta)$ we have that $\mathbb{P}[E] \geq 1-\eta$.

Consider a draw of v_{ij}'s that belongs to the event E. We have the following cases: For node $a \in A$, let $d_a(A)$ be the number of nodes a' such that $x_{aa'} = 1$. Also, let $d_a(B)$ be the number of nodes b such that $x_{ab} = 1$. Therefore, we have

$$u_a \leq (d_a^H(A) + d_a^H(B))(H-\gamma).$$

On the other hand, if all nodes in A such as a only connect to the nodes within their group, their utility would be

$$\begin{aligned}&(n-1)\left[r_a^H(A)(H-\gamma) + (1-r_a^H(A)(L-\gamma))\right] \\ &\quad\geq (n-1)(H-\gamma)((1-p-\delta)(1+a)-a),\end{aligned}$$

where we used $r_a^H(A) \geq (1-p-\delta)$ to obtain the last inequality where $a = \frac{\gamma - L}{H - \gamma}$. Since we have a strong Nash equilibrium, combining the last two relations, we obtain

$$(d_a^H(B) + d_a^H(A))(H-\gamma) \geq (n-1)(H-\gamma)((1-p-\delta)(1+a)-a).$$

We also have that $d_a^H(B) \leq n(1 - r_a^L(B)) \leq n(p+\delta)$. Using this inequality in the previous relation, we obtain

$$d_a^H(A)(H-\gamma) \geq (n-1)(H-\gamma)((1-p-\delta)(1+a)-a) - n(p+\delta)(H-\gamma),$$

which results in

$$d_a^H(A) \geq n\left((1-p-\delta)(1+a) - a - (p+\delta)\right).$$

For $\delta = 0$, since $p < \frac{1}{2+a}$, we have that $((1-p)(1+a) - a - p) > 0$. Therefore, for any choice of $p < \frac{1}{2+a}$, we can choose δ small enough such that

$$((1-p-\delta)(1+a) - a - (p+\delta)) > 0 \tag{A.4}$$

holds. Using lemma 4, there exist a strongly connected components in both groups A and B with size at least $n\left((1-p-\delta)(1+a) - a - (p+\delta)\right)$. Denote these two strongly connected components by S_A and S_B. Next, we show that no node from A-side is connected to S_B and similarly, no node from B-side is connected to S_A. For the sake of contradiction, suppose node b is connected to a node in S_A. Since S_A is strongly connected, b must be connected to all nodes in S_A. Next, we show that the utility b receives from connecting to all nodes of S_A is negative, resulting in a contradiction. The utility received from connecting to all nodes of S_A is upper-bounded by

$$n\left((1-p-\delta)(1+a) - a - (p+\delta) - (p-\delta)\right)(L-\gamma) + n(p-\delta)(H-\gamma)$$

For $\delta = 0$ if $p < \frac{1}{(a+1)^2+a}$, then this bound becomes negative, i.e.,

$$n\left((1-p-\delta)(1+a) - a - (p+\delta) - (p-\delta)\right)(L-\gamma) + n(p-\delta)(H-\gamma) < 0. \tag{A.5}$$

Therefore, one can choose δ small enough such that for $p < \frac{1}{(a+1)^2+a}$ this bound is negative. Another alternative for b is to only connect to highly popular nodes in S_A, which would give her a utility upper bounded by

$$-\gamma\beta n\left((1-p-\delta)(1+a) - a - (p+\delta) - (p-\delta)\right) + n(p-\delta)(H-\gamma).$$

For $p < \frac{1}{\frac{H-\gamma}{\gamma\beta}+a+3}$ and sufficiently small δ, this becomes negative, i.e.,

$$-\gamma\beta n\left((1-p-\delta)(1+a)-a-(p+\delta)-(p-\delta)\right)+n(p-\delta)(H-\gamma)<0 \quad \text{(A.6)}$$

Next, we show that all nodes in B are connected to all nodes of S_B. We show that the utility one receives from connecting to all nodes in S_B is positive. This utility is lower-bounded by

$$(n-1)\left((1-p-\delta)(1+a)-a-(p+\delta)-(p+\delta)\right)(H-\gamma)+(n-1)(p+\delta)(L-\gamma).$$

For $\delta = 0$ if $p < \frac{1}{2a+3}$, then this bound becomes positive. Therefore, one can choose δ small enough such that for $p < \frac{1}{2a+3}$ this bound is positive.

So far we proved that each group has a strongly connected component of size at-least

$$n\left((1-p-\delta)(1+a)-a-(p+\delta)\right)$$

denoted by S_A and S_B. All nodes in A are connected to S_A and all nodes in B are connected to S_B. Moreover, no node from A is connected to S_B and no node from B is connected to S_A. Next, we show that there is no connection from A to B and vice-versa. Suppose a node a from A is connected to a node in $B \backslash S_B$. Since all nodes of B are connected to S_B, a should connect to S_B as well (because his information is leaking to all nodes in S_B anyway). However, we show that this cannot happen in an equilibrium. Therefore, there is no connection between A and B. The overall bound on p is given by $p < \min\{\frac{1}{(a+1)^2+a}, \frac{1}{2a+3}, \frac{1}{\frac{H-\gamma}{\gamma\beta}+a+3}\}$. Given this inequality holds, one can choose δ small enough such all eq. (A.4), (A.5), and (A.6) hold simultaneously. This completes the proof.