Social Clubs and Social Networks

By Chaim Fershtman and Dotan Persitz*

We present a strategic network formation model based on membership in clubs. Individuals choose affiliations. The set of all memberships induces a weighted network where two individuals are directly connected if they share a club. Two individuals may also be indirectly connected using multiple memberships of third parties. Individuals gain from their position in the induced network and pay membership fees. We study the club congestion model where the weight of a link decreases with the size of the smallest shared club. A trade-off emerges between the size of clubs, the depreciation of indirect connections and the membership fee.
JEL: D71, D85, Z13.
Keywords: Strategic Formation of Clubs, Strategic Formation of Weighted Undirected Networks.

I. Introduction

Most of the initial social interactions between individuals occur within social circles, social groups or social clubs.[1] Clearly, some social connections can be formed randomly - like meeting someone on the street - but most friendships and acquaintances are formed within a social context like a family, school class, alumni organization, church, fraternity, academic department, research group, workplace, boy scouts, youth extracurricular activities, gym or even at a bar that the individual regularly attends.[2] That is, social links are typically formed within social contexts rather than between individuals who do not share any common social foci. Thus, when considering the formation of social networks, the social environment, particularly, the number and size of the different social clubs and the type of affiliations that people maintain within these clubs, also needs to be scrutinized.

* Coller School of Management, Tel Aviv University, Tel Aviv, 69978, ISRAEL. Emails: fersht@post.tau.ac.il and persitzd@post.tau.ac.il. We thank the editor and two anonymous referees for their insightful comments and suggestions. We benefited from conversations with Francis Bloch, Andrea Galeotti, Li Hao, Nils Röhl, Konrad Stahl and the participants of various workshops, conferences and seminars. We thank Zvika Messing, Amit Dekel and Omri Puny for excellent research assistance. Dotan Persitz acknowledges the financial support of the Henry Crown Institute of Business Research in Israel.

[1] Sociologists refer to social contexts as social foci. Feld (1981) introduces a "focus theory" where he defines social foci as "social, psychological, legal, or physical entities around which joint activities are organized." For the sociological literature see, Simmel (1908/1955), Young and Larson (1965*a*,*b*), Kadushin (1966), Feld (1981), Granovetter (1983), and the survey on non-geographical proximity by Rivera, Soderstrom and Uzzi (2010). See also the discussion on sub-neighborhoods in Jackson, Rodriguez-Barraquer and Tan (2012).

[2] In this paper we abstract from the direct benefits of belonging to a club. Our focus is on the role of clubs as platforms for the formation of social contacts. The direct benefits from belonging to a club are the focus of the well-established literature on local public goods (also referred to as club theory), starting with Tiebout (1956) and Buchanan (1965). Wooders (1978, 1980) was the first to introduce clubs into a general equilibrium framework to lay the foundations for a rigorous discussion on Tiebout (1956) hypotheses.

Most sociologists view social clubs as preceding the formation of social networks, as stated in Rivera, Soderstrom and Uzzi (2010, p. 106): "If networks are the fabric of inter-personal interaction, social foci are the looms in which they are woven." In some social clubs, membership is indeed automatic (e.g. family), but in most cases affiliation is by choice. People choose their gym, their university, their place of worship, as well as other social clubs that they wish to belong to, taking into account their existing club affiliations and the structure of their social network.[3]

We present a model where individuals choose affiliations in social clubs. The membership fee is identical across all clubs. Two individuals who share a club are linked in the induced social network. Thus, membership in a social club provides the benefit of being directly connected to other individuals in the club. Multiple club affiliations facilitate indirect connections. So, for example, an individual may have direct connections to her high school class mates in addition to having an indirect connection to an individual that attends a reading club together with one of her high school class mates.

Interaction in a small club is different than that in a large club. The "quality" of connection between two individuals generated in a large club, tends to be lower than that generated by a small club. In a small group, members are well acquainted and the flow of information is more reliable. Intuitively, the probability of any pair of members to interact and realize the potential benefit from their mutual affiliation decreases with the size of the club. To capture this congestion effect we assign each direct link with a weight which is a non-increasing function of the size of the smallest club shared by the two individuals. The weight of an indirect connection is the product of the weights associated with the links along the path. We define the shortest path between two individuals as the connection with the highest quality between them. That is, the shortest path between two individuals is the one that yields the highest product of the weights of the direct connections along the path. The benefits to an individual are the sum of the weights of the shortest paths to all other individuals net of the total club membership fees.

A social environment is Open Clubwise Stable if no individual wants to leave or join a club and there is no subset of individuals that are better off by forming a new club.[4] Open Clubwise Stability can be viewed as an extension of the pairwise stability solution concept posited by Jackson and Wolinsky (1996) to the club formation setup. That is, the connections model is equivalent to a club formation model with the restriction that clubs consist of exactly two members.[5]

Since Granovetter (1973) the concept of "weak ties" has become central to the applied literature on social networks. There are two possible interpretations of "weak ties." Two individuals may be connected directly via a large club that is subject to heavy congestion

[3]For sociological work that advocate for the simultaneous evolution of social networks and social foci see Feld (1981) and McPherson, Smith-Lovin and Cook (2001). Snijders et al. (2006) and Chandrasekhar and Jackson (2018) introduce stochastic non-strategic models of network formation that admit exogenously given clubs as platforms for link formation.

[4]While we highlight clubs as platforms on which links form, other works concentrate on the role of the coalition as a binding agreement that constrains player activities (Myerson (1980), Slikker and Van den Nouweland (2001), Caulier et al. (2013) and Caulier, Mauleon and Vannetelbosch (2015)).

[5]In Section 2 of the online appendix we show formally that the connections model with the pairwise stability solution concept is a special case of our framework with open clubwise stability.

or through an interconnected sequence of small clubs. In real-life both types of weak ties are observed and their relative importance depends upon the context. This paper highlights the trade-off between the two types of weak links - indirect connections composed of "high quality" links (the connection goes through a series of small clubs) and direct "low quality" links ("produced" in large clubs).[6]

In our club formation model, when the membership fee is sufficiently low, the environment wherein every pair of individuals share a single club of size two is the unique stable environment as the complete network is induced with high quality links. When the membership fee is higher, this environment is no longer sustainable and "weak ties" appear. We show that when congestion friction is higher than depreciation friction a stable environment is one based on "weak ties" of the indirect connections type. When depreciation is more significant than congestion, a stable environment is one based on "weak ties" generated in large clubs. In particular, our model predicts that complete networks can survive high membership costs if congestion is weaker than endogenous depreciation.

The trade-off between congestion and indirect connections is further demonstrated by considering the following two special environments: m-complete and m-star. In m-complete environments every pair of individuals shares exactly one club that includes m members, and therefore every pair of individuals is directly connected via a congested link (unless $m = 2$). In m-star environments, one individual (the "star") is affiliated with all the populated clubs, the other individuals (the "peripherals") are members of a single club and all the populated clubs are of size m. Therefore, in m-star environments, every peripheral individual is directly connected to $m-1$ individuals and indirectly connected to all the rest. We show that when the membership fee is low the efficient environment among the environments where all populated clubs are of size m, is the m-complete environment. The m-star environment is efficient for intermediate affiliation fees while the empty environment is efficient for sufficiently high membership costs.

We demonstrate that the stability of the various m-complete and m-star environments can be characterized as a function of the elasticity of congestion relative to club size. There is, however, non-monotonicity in the relationship between congestion and the size of clubs in stable environments. For a substantial set of congestion functions, m-complete environments with intermediate size clubs are never stable while m-complete environments with either small clubs (wherein each individual maintains many high quality affiliations) or large clubs (wherein each individual maintains few low quality affiliations) are open clubwise stable.

[6]Weak ties appear in two branches of the literature on networks in Economics. In the literature on the role of networks in labor markets, weak ties are typically viewed as cheap, infrequently used direct links that may relay useful job information (e.g. Montgomery (1992, 1994), Calvó-Armengol and Zenou (2003) and Kramarz and Skans (2014)). Calvó-Armengol (2004) studies job information transmission through indirect links but do not refer to those channels as weak ties. In the literature on the formation of weighted networks (which is frequently motivated by Granovetter (1973)) the weight is determined endogenously as some function of investments made by both end individuals (e.g. Goyal (2005), Brueckner (2006), Goyal, Konovalov and Moraga-González (2008) and Bloch and Dutta (2009)). Another approach is to model resource allocation as a subsequent stage to the formation of the network (e.g. Ballester, Calvó-Armengol and Zenou (2006) and Cabrales, Calvó-Armengol and Zenou (2011)). Altogether, this literature also interprets weak ties as direct links (with low weights) and does not refer to indirect connections as weak ties.

In real-life a wide range of rules regarding the formation, the joining or the leaving of social clubs are observed. For Example, clubs may introduce entry barriers wherein acceptance by incumbent members is required in order to join the club. Each set of rules induces a different set of possible deviations and therefore corresponds to a different stability concept. This clearly affects individual choices of clubs and consequently the stable environments. The solution concept of open clubwise stability represents an open environment wherein individuals are free to join or leave any club they wish, and to form new clubs, as long as they pay a fixed membership fee. To demonstrate the importance of club rules we introduce the Closed Clubwise Stability solution concept wherein joining a club requires the unanimous approval of all existing club members. We demonstrate that these two concepts may lead to dramatically different stable environments.

Establishing social connections via clubs yields different types of social networks than those formed in the regular framework of network formation. In particular, the club setting provides an alternative explanation for the extensive clustering that characterizes real-life social networks. In most real-life networks the probability of two individuals being connected if they are linked with a common individual is much higher than if the connections were formed randomly (for reviews see Goyal (2007) and Jackson (2008)). The social networks literature frequently attributes the high clustering to one of two explanations: First, individuals may have a preference for connections with individuals with whom they share a neighbor (preference for transitivity). Second, individuals may prefer to link to individuals with whom they share social traits (homophily). We argue that simultaneous formation of clubs and networks provides a third explanation for the high clustering observed in real-life networks, since each club is manifested as a clique in the induced social network.

II. The Model

An *environment* is a group of individuals and a set of clubs such that each individual is affiliated with a subset of clubs. Formally, $N = \{1,...,n_a\}$ ($n_a > 2$) is a finite set of individuals and $S = \{1,...,n_s\}$ is a finite set of clubs. The pair $\{i,s\}$ denotes the affiliation of Individual i with Club s and A^c is the set of all possible affiliations. An environment is the triplet $G \equiv < N,S,A >$ where $A \subseteq A^c$ is a set of affiliations. We denote the set of all the environments with n individuals by $\mathscr{G}_n$.[7] We denote the set of clubs that Individual i is affiliated with in Environment G by $S_G(i)$ and $s_G(i) \equiv |S_G(i)|$ denotes its cardinality. In addition, we denote by $N_G(s)$ the set of individuals that are affiliated with Club s in Environment G, and $n_G(s) \equiv |N_G(s)|$ denotes its cardinality. The environment that results from adding (severing) $\{i,s\}$ to (from) Environment G is denoted by $G+\{i,s\} \equiv < N,S,A \cup \left\{\{i,s\}\right\} >$ (similarly, $G-\{i,s\}$). Let $s \in S$ be a vacant club (we assume that such club always exists) and let $m \subseteq N$. Then, $G+m$ is the environment that

[7] A graph $G =< V,E >$ is called *bipartite* or *two mode network* if V admits a partition into two classes (U,$V \backslash U$) such that $\forall (v_1,v_2) \in E : v_1 \in U, v_2 \in V \backslash U$. An environment can be described as a bipartite graph wherein one set of nodes is the set of individuals and the other is the set of clubs. This representation is often referred to as an *affiliation network*. An additional way to represent an environment is by a hypergraph. A *hypergraph* is a pair $H =< U,ME >$ wherein the elements of ME (clubs) are subsets of U (individuals).

emerges from Environment G when the set m of individuals populates the vacant club s. Let c denote the club membership fee which is identical for all individuals in all clubs.[8]

Every Environment G induces an undirected network g whose nodes represent individuals and two individuals are linked in g if they belong to the same club. We denote the weight of a link between two individuals $i,i' \in N$ in G by $w(i,i',G) \in [0,1]$. In this general setting the weight can be a function of the whole environment and its interpretation is the quality of the link - the higher is the weight the "stronger" is the link.

Formally, a weighted network is a triplet $< N,E,W >$ wherein N is the set of individuals, E is the set of links and $W : E \to [0,1]$ the set of weights. The weighted network $g =< N,E_G,W_{G,w} >$ is induced by Environment G and weighting function w if $E_G \equiv \{\{i,j\}|i \in N, j \in N, S_G(i) \cap S_G(j) \neq \emptyset\}$ and $\forall \{i,j\} \in E_G : W_{G,w}(\{i,j\}) \equiv w(i,j,G)$. Note that each environment has another induced undirected network whose nodes represent clubs and two clubs are linked if there is an individual that affiliates with both. Fershtman and Gandal (2011) take advantage of this duality to study the open source environment.

We assume that individuals benefit from being connected, either directly or indirectly, to other individuals. Multiple affiliations facilitate indirect connections. Indirect connection between a pair of individuals occurs whenever a third party shares a club with each of the two individuals (see, for example, Faust (1997)). Formally, a *path* of length $l-1$ between Individual i and Individual i' in the induced network g is a sequence of individuals $\{x_1,x_2,\dots,x_{l-1},x_l\}$ such that $x_1 = i$ and $x_l = i'$ and every consecutive pair of individuals, x_k and x_{k+1}, shares at least one club in G. Two individuals who share at least one club are directly connected and two individuals who do not share a club in Environment G are indirectly connected if there is a path between them in g. If every pair of individuals is connected (either directly or indirectly) then g is connected; otherwise, it is disconnected. We say that Environment G is connected if its induced network g is connected.[9]

The *weight of a path* is the product of the weights on the links that constitute this path. That is, let $g =< N,E,W >$ be a weighted network. The weight of the path $p = \{x_1,\dots,x_l\}$ is $WP_g(p) = \prod_{k=1}^{l-1} W(\{x_k,x_{k+1}\})$. Path p is a *shortest weighted path* between individuals i and i' if and only if there is no path p' between individuals i and i' such that $WP_g(p') > WP_g(p)$. The *distance* between individuals i and i' in G using weighting function w, denoted $d(i,i'|G,w)$, is the weight of the shortest weighted path between them in the induced network g. If there is no such path, $d(i,i'|G,w) = 0$. Individuals benefit from short distances to other individuals. The utility of Individual i is given by $u_i(G,w,c) = \sum_{k \in N, k \neq i} d(i,k|G,w) - s_G(i) \times c$.

[8]For simplicity we assume that the membership fee is fix. More generally, one may consider a case where the club membership fee depends on the size of the club or its composition (e.g. Buchanan (1965), McGuire (1974), Wooders (1988, 1989), Haller (2016)).

[9]Consider two environments $G =< N,S,A >$ and $G' =< N',S',A' >$. If $S' \subseteq S$, $N' = \cup_{s \in S'} N_G(s)$ and $A' = \{\{i,s\}|i \in N', s \in S', \{i,s\} \in A\}$ then G' is a sub environment of G and G is a super environment of G'. If, in addition, $N' = N$ then G' is a spanning sub environment of G and G is a spanning super environment of G'. The sub environment $G' =< N',S',A' >$ of $G =< N,S,A >$ is a component of G if its induced network g' is connected and there is no pair of individuals $i \in N'$ and $k \in N \backslash N'$ who share a club in G. We denote the size of the component G' by $n(G') = |N'|$.

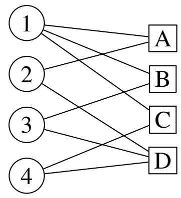
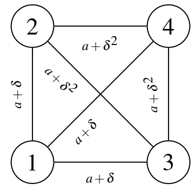
Environment	List of Clubs	Induced Network	Utilities
 	Club A: 1 2 Club B: 1 3 Club C: 1 4 Club D: 2 3 4	 	$u_1 = 3(a+\delta) - 3c$ $\forall i \in \{2,3,4\}$: if $\delta \geq \frac{1-a}{2}$: $u_i = (a+\delta) + 2(a+\delta)^2 - 2c$ Otherwise: $u_i = (a+\delta) + 2(a+\delta^2) - 2c$

FIGURE 1. AN ENVIRONMENT, ITS INDUCED WEIGHTED SOCIAL NETWORK AND THE INDIVIDUALS' UTILITIES.

Figure 1 provides a simple example. The two leftmost cells describe an environment containing four individuals and four populated clubs wherein Individual 1 shares a club of size two with each of the other three individuals who, among themselves share an additional club of size three. Therefore, the induced weighted network is the complete network depicted in the next cell. The weights on the links are such that a club of size two provides a link of strength $a+\delta$ while a club of size three provides a weaker link of strength $a+\delta^2$ ($\delta \in (0,1)$, $a \in [0,1)$ and $a+\delta \in (0,1)$). Finally, the utilities of the individuals are documented in the rightmost column. As this example demonstrates the shortest distance is not necessarily the path that includes the least number of links.

Environment G is *Open Clubwise Stable* (henceforth, OCS) if no individual strictly gains from leaving a club, no individual strictly gains from joining a club and there is no subset of individuals who are all better off by forming a new club together. Formally, the conditions for OCS are:

(i) No Leaving: $\forall s \in S, \forall i \in N_G(s) : u_i(G,w,c) \geq u_i(G-\{i,s\},w,c)$.

(ii) No New Club Formation: $\forall m \subseteq N$:
$\exists i \in m : u_i(G+m,w,c) > u_i(G,w,c) \Rightarrow \exists j \in m : u_j(G+m,w,c) < u_j(G,w,c)$.

(iii) No Joining: $\forall s \in S, \forall i \notin N_G(s) : u_i(G,w,c) \geq u_i(G+\{i,s\},w,c)$.

Environment G is *Strongly Efficient* (henceforth, SE) if there is no other environment G' such that $\sum_{i\in N} u_i(G',w,c) > \sum_{i\in N} u_i(G,w,c)$.

The following environments are instrumental in characterizing stable and efficient environments in this setting.

(i) $G = < N,S,\emptyset >$ is the Empty environment.

(ii) G is the Grand Club environment if there is exactly one populated club and all the individuals are affiliated with it.

(iii) G is the All Paired environment if every pair of individuals shares a unique club of size two.

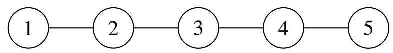
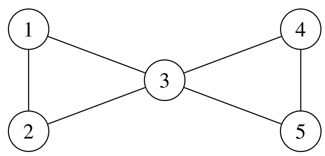
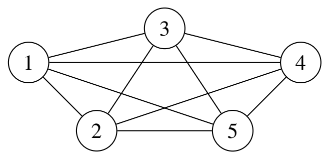
Environment	Induced Network	Class ("Weakest Affiliation")
Club A: 1 2 Club B: 2 3 Club C: 3 4 Club D: 4 5		K(G)=1 Individual 2 leaves Club A Individual 4 leaves Club D
Club A: 1 2 3 Club B: 3 4 5		K(G)=2 Individual 3 leaves Club A Individual 3 leaves Club B
Club A: 1 2 3 4 5		K(G)=4 Every individual that leaves Club A

FIGURE 2. THREE MINIMALLY CONNECTED ENVIRONMENTS OF 5 INDIVIDUALS, THEIR INDUCED NETWORKS AND THEIR CLASSES.

III. Baseline Model: No Congestion

We start by considering the simple setting in which all weights are set to 1 ($w(i,j,G)$ is identically 1), implying that the distance between two individuals is 1 if they are (either directly or indirectly) connected and 0 otherwise.

When membership is free, individuals wish to maximize the number of other individuals with whom they are connected, either directly or indirectly. In this case it is hardly surprising that every connected environment is both stable and efficient.

For the case of a positive membership fee, we say that environment G is *Minimally Connected* if it is connected and for every affiliation $\{i,s\} \in A$, the network induced by $G - \{i,s\}$ is disconnected. In fact, it is easy to show that $G - \{i,s\}$ contains exactly two components - one that contains Individual i, denoted $C_i(G - \{i,s\})$, and one that does not contain Individual i, denoted $C_{-i}(G - \{i,s\})$. In the setting with no congestion, the size of $C_{-i}(G - \{i,s\})$ is the loss incurred by Individual i upon canceling the affiliation with Club s. We say that the "weakest affiliation" in Environment G is the one whose absence leads to the smallest $C_{-i}(G - \{i,s\})$. We classify the minimally connected environments by their "weakest affiliation," defined by $K(G) = \min_{\{i,s\} \in A} n(C_{-i}(G - \{i,s\}))$ where $n(C)$ is the number of individuals in component C. Figure 2 demonstrates this classification on some minimally connected environments that contain 5 individuals.

Proposition 1. *Suppose that for every environment G and for every pair of individuals i and j who share a club in G,* $w(i, j, G) = 1$. *(i) When* $n_a - 1 > c > 0$, *(a) G is a Minimally Connected environment of class* $K(G) \geq c$ *if and only if G is OCS and (b) The Grand Club is the unique SE environment. (ii) When* $c > n_a - 1$, *the Empty environment is the unique OCS and SE environment.*

All proofs are relegated to Appendix A of the online Appendix. The intuition behind Proposition 1 is: First note that for $n_a - 1 > c > 0$ the Grand Club environment is OCS while the Empty environment is not. Hence, if G is OCS and disconnected there must be a component H that contains $n_a > h > 1$ individuals. Since the maximal possible utility of an individual in H is $(h-1) - c$ and since G is OCS then $c < h - 1$. But then it is beneficial for every individual who is not included in H to join any one of H's clubs. Therefore, if G is OCS then it is connected. But, if it is not minimally connected there is an individual who may want to leave a club since leaving will not affect network connectivity (i.e. the individual's benefits). Finally, if the membership costs are higher than the value of the "weakest affiliation" there will be individuals who can improve by canceling one of their affiliations.[10]

IV. The Club Congestion Model

The quality of the connections generated within a club may depend on the size of the club. If many individuals are affiliated with the club, the "quality" of the connection between any two members is probably lower than the "quality" of the connection between any two members of a small club. For example, consider the difference between belonging to a club of five individuals who attended the same college together versus being a member of the club of the class of '87 at a high school with over two hundreds members. In this model we capture "quality" by assuming that links are weighted and that the weight of each link depends upon the size of the club shared by the individuals. Specifically, ***The club congestion function*** is a non-increasing function $h : \{2, 3, \ldots, n_a\} \to [0, 1]$. Furthermore, we assume that when individuals share more than one club, the weight of the link between them is determined by the congestion in the smallest club that they share. Formally, given club congestion function h, the weight of a link between two individuals $i, i' \in N$ is $w_h(i, i', G) = \max_{s \in S_G(i) \cap S_G(j)} h(n_G(s))$.

Even without congestion the affiliations of one individual may affect the social network of other individuals. Unilateral actions such as leaving a club or joining a club may benefit or harm other individuals by creating new links (either direct or indirect) or by "breaking" some of the shortest paths. Incorporating congestion into the club formation setting introduces a new type of externality whereby these unilateral actions may also affect the quality of some links. For example, if Individual j joins a club with which Individual i is also affiliated, the quality of some links that Individual i maintains may

[10] Bar (2005) also considers a model of strategic formation that includes club structure. However, she ignores any type of congestion and therefore her results are comparable to our Proposition 1 with some minor differences. See also Jun and Kim (2009), Borgs et al. (2011) and So, Mui and Rai (2017).

change - either by making some paths shorter or by reducing the weight of some links due to stronger congestion. While this externality does not affect an individual's decision either to join or leave a club (Individual j in the example) it clearly affects the social desirability of the new environment. Importantly, while unilateral actions may have positive or negative externalities, the formation of a new club can never hurt uninvolved individuals. Therefore, if Environment G is Strongly Efficient it must satisfy the condition of "No New Club Formation".

Our analysis of club congestion begins by examining the simple case in which membership is free. Clearly, if $1 > h(2) > h(3)$, the only OCS environments are the spanning super environments of the All Paired environment and these are also the only efficient environments. That is, in any efficient OCS environment every pair of individuals must share a club of size two. Therefore, the induced network is complete and the weights on the links are the highest possible since club congestion is at its minimum.

We now turn to consider the club congestion model with positive membership fee. We begin with some preliminary results and necessary definitions, then in Section IV.A we move to introduce two families of architectures - the m-Complete environments and the m-Star environments. Section IV.B provides a result on the efficient architectures in the club congestion model. Sections IV.C and IV.D analyse the open clubwise stability of the m-Complete environments and the m-Star environments. Section IV.E uses these results to provide a novel insight on the nature of weak links.

For our first result note that an individual holds at most $n_a - 1$ links in the induced network and this network includes at most $\frac{n_a(n_a-1)}{2}$ links. Since each link is determined by a single club - the smallest club the two individuals share - we can establish an upper bound on the number of affiliations per individual and the number of populated clubs in an OCS environment assuming affiliations are costly.

Lemma 1. *Suppose $c > 0$. If Environment $G =< N,S,A >$ is OCS then:*

(i) $\forall i \in N : s_G(i) \leq n_a - 1$.

(ii) $|\{s \in S | n_G(s) > 0\}| \leq \frac{n_a(n_a-1)}{2}$.

The number of possible clubs in an environment with n_a individuals is $2^{n_a} - (n_a + 1)$. Therefore, Lemma 1.(ii) implies that OCS environments in the club congestion model include relatively few populated clubs.[11]

The set of OCS environments in the club congestion model with positive membership fee crucially depends on the properties of the congestion function. We therefore introduce two functional forms of club congestion, Reciprocal Club Congestion and Exponential Club Congestion. These will be useful in demonstrating some of the results in the upcoming analysis.

Reciprocal Club Congestion: Environment G is characterized by Reciprocal Club Congestion if $\forall m \geq 2 : h(m) = \frac{1}{m-1}$.

[11] E.g. for 10 individuals there are 1013 possible clubs, but an OCS environment includes at most 45 populated clubs. We use Lemma 1 in the Matlab code package that accompanies this work (the package can be found on GitHub: `https://github.com/omri1348/Social-Clubs-and-Social-Networks/tree/master/code`).

Exponential Club Congestion: Environment G is characterized by Exponential Club Congestion if $h(m) = a + \delta^{m-1}$ where $\delta \in (0,1)$, $a \in [0,1)$ and $a + \delta \in (0,1)$.

The Reciprocal Club Congestion function can be interpreted as one unit of attention that individuals in a club uniformly lavish upon the other club members. The Exponential Club Congestion function is the sum of two components: The first, representing the role of the club as an institution that connects individuals is a constant denoted by a, and therefore depends only on individuals' mutual affiliation and the second, that can be interpreted as the prospects of a potential link materializing, is an exponential function that decreases with the size of the club, δ^{m-1} ($\delta \in (0,1)$).

When individuals are affiliated with a club of size m they enjoy $m-1$ direct links to other club members. We define $k_h(m) \equiv (m-1) \times h(m)$ as the Direct Club Value (henceforth, DCV). The size of club, m, has two effects on $k_h(m)$. While a bigger club generates more direct connections, these links are of lower quality due to club congestion.

The reciprocal club congestion function is a special case of the two effects of club size on the DCV cancelling each other out as $k_h(m)$ is equal to 1 independently of m. Intuitively, for a club member, the direct value of a club is exactly the unit of attention collected from other members. The DCV of the exponential congestion function depends on a and δ. When $a = 0$, it can be shown that when $\delta < \frac{1}{2}$ the congestion effect is dominant and the DCV is maximized when the club is small ($m = 2$), but when $\delta > \frac{1}{2}$ a higher value of δ implies that the DCV is maximized by a larger value of m. When $a > 0$, the effect of the number of links is reinforced since the aggregate benefit of a increases linearly with m.

To demonstrate the role of the DCV consider the Empty environment. The Empty environment always satisfies both the conditions of "No Leaving" and "No Joining". Therefore, the Empty environment is OCS if and only if the condition of "No New Club Formation" holds (hence if the Empty environment is not OCS it is also inefficient). Notice that the benefit of an individual from participating in the formation of a new club of size m is exactly the DCV of this club, $k_h(m)$. Therefore, the Empty environment with n_a individuals is OCS if and only if $c \geq \max_{m \in \{2,\dots,n_a\}} k_h(m)$. For a detailed discussion on the DCV of the exponential congestion function and on the stability of the Empty environment see Section 3 in the online appendix.

Lemma 2 below connects the strategic aspects captured by the DCV to the properties of the club congestion function. We define the club-size elasticity of the club congestion function h as $\eta_h(m) \equiv \frac{\frac{h(m+1)-h(m)}{h(m)}}{\frac{1}{m}}$ for every club size m where $h(m) > 0$ and $\eta_h(m) \equiv 0$ otherwise. $h(m)$ is non-negative and non-increasing and, therefore, $\eta_h(m) \leq 0$. We say that $h(m)$ is inelastic (elastic) if $\forall m \in \{2,\dots,n_a - 1\} : \eta_h(m) > -1$ (respectively, $\eta_h(m) < -1$).

Lemma 2. *The club congestion function $h(m)$ is inelastic (elastic) if and only if $k_h(m)$ is strictly increasing (decreasing).*

A. Two Useful Clubs Architectures

The next step of our analysis is to introduce two new general architectures that play a central role in characterizing the stable and efficient environments in the Club Congestion Model.

Note that the architectures introduced so far exist for any number of individuals. However, the two types of architectures defined below may not exist for some sizes of the society due to the familiar integer problems. Such problems, in the context of equilibrium existence, were discussed first in the local public good literature, starting with Wooders (1978, 1980), and then in the network formation literature, starting with Page and Wooders (2007) and Arnold and Wooders (2015). While several ingenious solution concepts that deal with this problem are provided in this literature (e.g. the ergodic club equilibrium of Arnold and Wooders (2015)), for brevity, in this work we ignore such integer issues.

THE m-COMPLETE ENVIRONMENT. — G is an **m-Complete environment** ($m \in \mathbb{N}$, $n_a \geq m \geq 2$) if $\forall i, i' \in N : |S_G(i) \cap S_G(i')| = 1$ and $\forall s \in S : n_G(s) = m$ or $n_G(s) = 0$. That is, in m-Complete environments every pair of individuals shares exactly one club and all the populated clubs are of the same size, m.[12]

Figure 3 provides two examples of m-complete environments. The first example exhibits the All Paired environment. Compared to other m-complete environments, the links are stronger but an individual needs to join more clubs in order to be connected to all the other individuals. To demonstrate this trade-off consider the case of the 3-Complete environment with seven individuals. In this environment each individual is a member of three clubs while in the All Paired environment with $n_a = 7$ each individual pays for six memberships. At the same time, the links in the network induced by the All Paired environment are stronger than those in the network induced by the 3-Complete environment.

Another important observation is that in m-complete environments indirect connections are never the shortest paths since every pair of individuals is connected by a direct link and all links are identically weighted. While most of the other environments contain frictions due both to congestion and indirect connections, m-complete environments are free of the friction caused by indirect connections.

THE m-STAR ENVIRONMENT. — In the literature on the strategic formation of social networks the star network often emerges as both stable and efficient for medium levels of linking costs. The star structure has one individual who maintains links with all the other individuals while these individuals have no additional direct connections. We generalize this topology by defining the m-Star environment, where the size of the clubs is $m \geq 2$,

[12] Given n_a and m, a necessary condition for the existence of an m-Complete environment is that $\frac{n_a-1}{m-1}$ and $\frac{n_a(n_a-1)}{m(m-1)}$ are integers. As a combinatorial object an m-Complete environment with n_a individuals is the "Steiner System" $S(t, m, n_a)$ where $t = 2$.

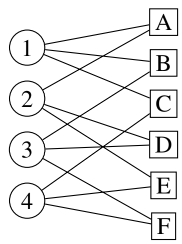
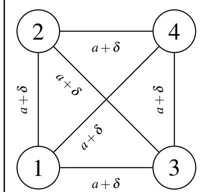
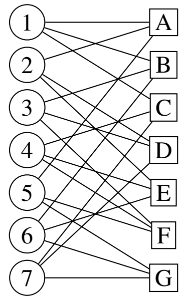
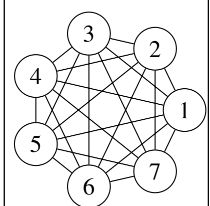
	Environment	List of Clubs	Induced Network	Utilities
All Paired or 2-Complete (n=4)		Club A: 1 2 Club B: 1 3 Club C: 1 4 Club D: 2 3 Club E: 2 4 Club F: 3 4		$\forall i \in \{1,2,3,4\}$**:** $u_i = 3(a+\delta) - 3c$
3-Complete (n=7)		Club A: 1 2 5 Club B: 1 3 6 Club C: 1 4 7 Club D: 2 3 7 Club E: 2 4 6 Club F: 3 4 5 Club G: 5 6 7	 The weights are all $a+\delta^2$	$\forall i \in \{1,\ldots,7\}$**:** $u_i = 6(a+\delta^2) - 3c$

FIGURE 3. TWO m-COMPLETE ENVIRONMENTS, THEIR INDUCED WEIGHTED NETWORKS (WEIGHTED BY THE EXPONENTIAL CLUB CONGESTION FUNCTION) AND THE INDIVIDUALS' UTILITIES.

one individual is a member of all clubs and all other individuals are members of a single club.[13] Formally,

m-Star: G is an *m-Star environment* ($m \in \mathbb{N}$, $n_a \geq m \geq 2$) if:

$$(i) \quad \forall s \in S : n_G(s) = m \text{ or } n_G(s) = 0.$$

$$(ii) \quad \exists i \in N \text{ such that } \forall s', s'' \in \{s | n_G(s) > 0\} : N_G(s') \cap N_G(s'') = \{i\}.$$

$$(iii) \quad \forall j \in N \backslash \{i\} : s_G(j) = 1.$$

Two m-Star environments are demonstrated in Figure 4. In 2-Star environments there is one individual who is a member of $n_a - 1$ clubs of size two with all the other individuals and therefore provides all the connectivity in the induced network. Each club, in this example, induces a weight of $a+\delta$ and the distance between each pair of these $n_a - 1$ individuals is $(a+\delta)^2$.

In the 3-Star environment, all clubs are of size three and all include one special individual. Compared to the 2-Star environment the central individual in the 3-Star environment pays lower membership fees but suffers greater congestion. The larger m, the more direct

[13]Given n_a and m, a necessary and sufficient condition for the existence of an m-star environment is that $\frac{n_a-1}{m-1}$ is an integer. Note that these networks are special cases of the $m-1$-quilts introduced by Jackson, Rodriguez-Barraquer and Tan (2012).

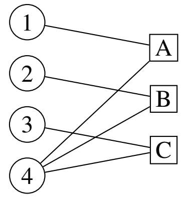
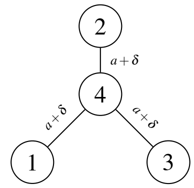
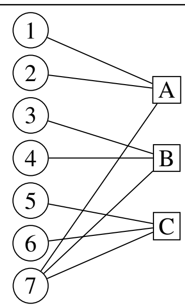
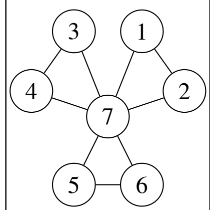
	Environment	List of Clubs	Induced Network	Utilities
2-Star (n=4)		Club A: 1 4 Club B: 2 4 Club C: 3 4		$\forall i \in \{1,2,3\}$: $u_i = (a+\delta) + 2(a+\delta)^2 - c$ $u_4 = 3(a+\delta) - 3c$
3-Star (n=7)		Club A: 1 2 7 Club B: 3 4 7 Club C: 5 6 7	The weights are all $a+\delta^2$	$\forall i \in \{1,\dots,6\}$: $u_i = 2(a+\delta^2) + 4(a+\delta^2)^2 - c$ $u_7 = 6(a+\delta^2) - 3c$

FIGURE 4. TWO m-STAR ENVIRONMENTS, THEIR INDUCED WEIGHTED NETWORKS (WEIGHTED BY THE EXPONENTIAL CLUB CONGESTION FUNCTION) AND THE INDIVIDUALS' UTILITIES.

connections peripheral individuals have but the lower the quality of their connections, both direct and indirect. Thus, while individuals in m-Complete environments suffer only from congestion and individuals in 2-Star environments suffer only from depreciation caused by their indirect connections, individuals in m-Star environments generally suffer from both types of frictions.

B. Efficiency

In many standard homogeneous models of strategic network formation (e.g. the connections model of Jackson and Wolinsky (1996)), strongly efficient topologies are the complete network for low linking costs, the star network for medium linking costs and the empty network for high linking costs. These results reflect the benefits of direct linking and the role of short indirect connections as a substitute for direct connections when linking costs are substantial.

Proposition 2 demonstrates that a similar intuition pertains in the Club Congestion Model with respect to constant levels of congestion. In order to control for the level of congestion friction we consider the set of m-Uniform environments in which all populated clubs are of size m. That is, G is an m-Uniform Environment ($m \in \{2,\dots,n_a\}$) if $\forall s \in S : n_G(s) = m$ or $n_G(s) = 0$. Denote the set of all m-Uniform environments with n individuals by $\mathscr{G}_n^m$ and denote the set of all uniform environments with n individuals by $\mathscr{G}_n^{all} = \cup_{k=2}^{n_a} \mathscr{G}_n^k$. Proposition 2 implies that the strongly efficient uniform environments are m-Complete, m-Star, or Empty.

Proposition 2. *Let $m \in \{2, \ldots, n_a\}$. For every club congestion function $h(\cdot)$ and m-Uniform Environment $G' \in \mathscr{G}_n^m$:*

(i) Let $c \in [0, (m-1)(h(m)-h^2(m))]$ and let G be an m-Complete environment. Then, $\sum_{i=1}^{n_a} u_i(G, w_h, c) \geq \sum_{i=1}^{n_a} u_i(G', w_h, c)$.

(ii) Let $c \in [(m-1)[h(m)-h^2(m)], (m-1)h(m) + \frac{(n_a-m)(m-1)}{m} h^2(m)]$ and let G be an m-Star environment. Then, $\sum_{i=1}^{n_a} u_i(G, w_h, c) \geq \sum_{i=1}^{n_a} u_i(G', w_h, c)$.

(iii) Let $c \geq (m-1)h(m) + \frac{(n_a-m)(m-1)}{m} h^2(m)$ and let G be the Empty environment. Then,
$\sum_{i=1}^{n_a} u_i(G, w_h, c) \geq \sum_{i=1}^{n_a} u_i(G', w_h, c)$.

Proposition 2 considers the set of environments in which all clubs are of size m. When the membership fee is low, the m-Complete environments are efficient. Since m-Complete environments are symmetric across individuals, the upper bound is independent of n_a and represents the individuals' preference for costly direct links ($(m-1)h(m)-c$) over free indirect links ($(m-1)h^2(m)$). When the membership fee increases, the importance of short indirect connections relative to costly direct connections and the low quality of long indirect connections emerges. The architecture of m-Star environments implements these preferences since the direct connections of the central individual keep the environment connected while making all other connections as short as possible. While the lower bound is independent of n_a, the upper bound increases with n_a since the larger the environment, the larger the return for membership for everyone except the central individual.

In the proof we first show that when $c \leq (m-1)(h(m)-h^2(m))$, the m-Complete environment achieves maximal total utility among all connected m-Uniform environments with no more than $\frac{n_a(n_a-1)}{m(m-1)}$ clubs due to the high quality of the direct connections. This result also holds when the m-Complete environment is compared to large connected m-Uniform environments (since additional clubs are redundant) and to disconnected m-Uniform environments (since the total utility of the m-Complete environment is convex in n_a). A result on hypergraphs from Berge (1989) is adopted to show that m-Star environments minimize the number of clubs required for an m-Uniform environment to be connected. When $c \geq (m-1)(h(m)-h^2(m))$, the m-Star environment achieves the maximal total utility among all connected m-Uniform environments due to the high quality of the indirect connections and the low total membership fees. This result also holds when the m-Star environment is compared to non-empty disconnected m-Uniform environments since the union of two stars has a higher total utility than the sum of the totals of the two stars (due to additional indirect connections). The third part results from the fact that when $c > (m-1)h(m) + \frac{(n_a-m)(m-1)}{m} h^2(m)$, the total utility of the m-Star environment is negative.

The comparison of efficient uniform environments across club sizes depends on the specific congestion function. However, two implications can be drawn from Proposition 2. One is that when the membership fee is below $\min_{m \in \{2,\ldots,n_a\}} (m-1)(h(m)-h^2(m))$, the environment that achieves the maximal total utility among all uniform environments must be an m-Complete environment. The other is that since the maximal DCV across club sizes is greater than $\max_{m\in\{2,\dots,n_a\}}(m-1)(h(m)-h^2(m))$, among all uniform environments only an m-Star environment or the Empty environment may achieve maximal total utility when the membership fee is higher than the maximal DCV across club sizes.

C. *The Stability of m-Complete Environments*

The m-complete environment is Open Clubwise Stable only when the membership fee is neither too high nor too low. Low membership fee is an incentive to form new small clubs (if $m > 2$) while high membership fee is an incentive to leave one of the clubs, replacing direct connections with indirect ones. Note that in the m-Complete environment, the "No Joining" condition is irrelevant since if an individual joins an existing club, that individual pays additional membership fee but creates no new (or better) connections.

Proposition 3. *Let $\hat{k}$ denote the club size that maximizes the Direct Club Value and let $n_a > m \geq 2$. An m-Complete environment is OCS if and only if*

$$c \in \Big[\max_{k\in\{2,\dots,\min\{m-1,\hat{k}\}\}}(k-1)[h(k)-h(m)],(m-1)[h(m)-h^2(m)]\Big]$$

In m-Complete environments, forming new clubs of size smaller than m may reduce congestion. Hence, the lower bound of Proposition 3 implies that a necessary condition for an m-Complete environment to be OCS is that the membership fee is high enough to preclude new clubs from being formed. The benefit of a coalitional deviation to a club of size $k < m$ is its DCV ($(k-1)h(k)$) net the value of these links in the original environment ($(k-1)h(m)$). The DCV applies here since indirect connections never constitute the shortest path for new club formation deviations in m-Complete environments. Note that the larger the new club, the larger the number of original links whose value has been lost. Hence, deviation to clubs with more than $\hat{k}$ members is less attractive than deviation to clubs of size $\hat{k}$ because of the lower DCV and the greater loss of original value.

An individual may consider leaving a club to trade-off reduced membership payments with replacing some direct connections with indirect ones (all individuals maintain multiple memberships, $n_a > m$). Proposition 3 guarantees that the membership fee is not high enough to make such a trade-off worthwhile. Note that the existence of a stable m-Complete environment is not guaranteed. It is possible that the lower bound is higher than the upper bound (see further discussion in Section 4 of the online appendix).

We now focus on two extreme cases which are of special interest - the All Paired and the Grand Club environments. The stability of these environments depends on the relative importance of the two frictions - club congestion and the membership fee. While in the All Paired environment individuals suffer high membership fees ($n-1$ clubs) but no club congestion, in the Grand Club environment congestion is strong but membership fees are minimal.

THE STABILITY OF THE ALL PAIRED ENVIRONMENT. — In the All Paired environment the individuals suffer no congestion and no depreciation. When membership fee is introduced the strict super environments of the All Paired are no longer OCS since redundancy is costly. In the All Paired environment, joining an existing club or forming a new one never constitutes a beneficial deviation since additional affiliations are costly and individuals already share small clubs with everyone else. Therefore, only incentives to leave a club and use an indirect connection instead are relevant. As long as the gain from a direct connection ($h(2)-c$) is greater than an indirect connection ($h^2(2)$), the All Paired environment is OCS.

This argument, however, does not rule out the stability of environments wherein the smallest club shared by some pair of individuals is larger than size 2. Such environments are not OCS if the cost of forming a new club of size 2 is lower than the benefit derived from eliminating the club congestion suffered by this pair, that is, when $h(2)-h(3)>c$. Therefore, the uniqueness of the All Paired environment is guaranteed when the membership fee is small enough to allow individuals to form new two-individual clubs in order to resolve both the friction created by indirect connections and the friction of club congestion. Formally, the All Paired environment is the unique OCS environment when $c \in (0, \min\{h(2)-h^2(2), h(2)-h(3)\})$. In fact, in this range, the All Paired environment is also the unique SE.

THE STABILITY OF THE GRAND CLUB ENVIRONMENT. — In the Grand Club environment individuals suffer severe club congestion but no depreciation friction and minimal membership fees. Proposition 1 states that when there is no club congestion and $n_a-1>c>0$ the Grand Club environment is both OCS and the unique efficient environment. Claim 1(i) provides a necessary and sufficient condition for the existence of OCS Grand Club environment when congestion exists.

Claim 1. *Denote by $\hat{k}$ the club size that maximizes the Direct Club Value.*

(i) The Grand Club environment is OCS if and only if

$$c \in \Big[\max_{k\in\{2,\dots,\min\{n_a-1,\hat{k}\}\}} (k-1)[h(k)-h(n_a)], (n_a-1)h(n_a)\Big]$$

(ii) If the club congestion function is inelastic, a range of membership fees in which the Grand Club environment is OCS exists.

(iii) Let $n_a \geq 4$ and let $h(\cdot)$ be an exponential club congestion function where $\delta \in (0, \frac{1}{2})$. For $a=0$ the Grand Club environment is never OCS. But, if $a>0$, there exists an $\bar{n}_a$ such that $\forall n_a : n_a > \bar{n}_a$, a range of membership fees in which the Grand Club environment is OCS exists.

Since the DCV of the reciprocal club congestion function is unity, the most attractive deviation from the Grand Club environment is to a club of size 2 (wherein the loss of original value is minimal). Therefore, when club congestion is reciprocal, the Grand Club environment is OCS if and only if $c \in [1-\frac{1}{n_a-1}, 1]$. Hence, when the club congestion function is reciprocal a range of membership fees in which the Grand Club environment is OCS always exists. Generally, however, Claim 1(i) does not guarantee that such a range exists. Nevertheless, Lemma 2 is used to demonstrates that when club congestion is not too sensitive to club size, the Grand Club environment is OCS for some range of membership fees (Claim 1.(ii)).

When the club congestion function is exponential, Claim 1.(ii) guarantees that the Grand Club environment is OCS for some range of membership fees in some cases (e.g. when $a = 0$ and $\delta > 1-\frac{1}{n_a}$). The case of $\delta \in (0, \frac{1}{2})$ wherein the congestion component of the club congestion function is substantial is analyzed in Claim 1.(iii). The first part demonstrates that when there is no non-congested component, congestion is too strong for a Grand Club environment to be OCS. However, if there is some non-congested component, the Grand Club environment can be OCS for some range of membership fees, as long as the set of individuals is large enough to make the non-congested part important. A sociological interpretation may imply that social solidarity (which does not depend on club size) may be useful in maintaining big clubs even when club congestion is strong.

D. The Stability of m-Star Environments

An *m*-Star environment is OCS if no individual wishes to join or leave an existing club and no subset of individuals benefits from forming a new club.

The central individual prefers to leave a club when the membership fee is higher than the benefit derived from direct links to the individuals in the club. A peripheral individual wishes to leave a club when the membership fee is higher than the benefit of direct links to other club members and indirect connections to all other peripheral individuals. Therefore, the peripheral individuals' incentives to leave a club are weaker than those of the central individual. Thus, the upper bound on the range of membership fees in which an *m*-Star environment is OCS depends on the central individual's incentives. Since by leaving a club, the central individual disconnects from the other members of the club, the upper bound is higher than in *m*-Complete environments where direct links that have been lost can be replaced by indirect connections.

Joining an existing populated club is not a relevant consideration for the central individual since this individual is already a member of all populated clubs. When joining an existing club, a peripheral individual replaces $m-1$ indirect connections with costly and congested direct connections (relative to existing direct connections). Thus, the lower bound on the range of membership fees in which an *m*-Star environment is OCS should be high enough to make the existing indirect connections more attractive than new direct connections for a peripheral individual.

The third consideration is the formation of a new club. In forming a new club, a peripheral individual always gains more than the central individual. If the new club is smaller than *m*, the central individual only gains from improved direct connections while peripheral individuals also gain from better indirect connections. When the new club is weakly larger than *m*, the central individual gains nothing while peripheral individuals may gain from the new direct links created. Therefore, the lower bound on the range of membership fees for which an m-Star environment is OCS should be also high enough to deter peripheral individuals from forming new clubs.

A peripheral individual always prefers to form a new club with members of other clubs. If the new club is of size $k < m$, sharing it with an individual who is also affiliated with the original club yields a single improved direct connection $(h(k) - h(m))$ while forming the new club with an individual who belongs to a different original club turns an indirect connection into a direct one $(h(k) - h^2(m))$ and it may also have a positive effect upon other indirect connections. If the new club is not smaller than m, then, sharing this new club with an individual who is also affiliated with the original club yields nothing while forming a new club with an individual who belongs to a different original club may improve one indirect connection. Thus, the attractiveness of a new club increases with the number of original clubs that are represented in it. Hence, the lower bound on the range of membership fees in which an m-Star environment is OCS should be high enough to deter peripheral individuals from coordinating the formation of a new club that includes a diverse collection of members relative to the original clubs. Proposition 4 summarizes these incentives.

Proposition 4. *Let $n_a > m \geq 2$ and let $h(\cdot)$ be the club congestion function. Denote $\gamma \equiv \frac{n_a-1}{m-1}$, $\eta_k \equiv \lceil \frac{k}{\gamma} \rceil$ and $l_h = \min\{k \in \mathbb{Z} | h(k) \leq h^2(m)\}$.*

(i) If $\gamma \geq m$ the m-Star environment is OCS if and only if

$$k_h(m) \geq c \geq \max\{\max_{m \geq k \geq 2} FNS_h(k,m), \max_{\min\{l_h,n_a\} > k > m} FNL_h(k,m,n_a)\}$$

(ii) If $\gamma < m$ the m-Star environment is OCS if and only if[14]

$$k_h(m) \geq c \geq \max\{J_h(m), \max_{\gamma \geq k \geq 2} FNS_h(k,m), \max_{m \geq k > \gamma} FNI_h(k,m,n_a), \max_{\min\{l_h,n_a\} > k > m} FNL_h(k,m,n_a)\}$$

where

$$FNS_h(k,m) = (k-1)[h(k) + (m-2)h(k)h(m) - (m-1)h^2(m)]$$

$$FNL_h(k,m,n_a) = (k-\eta_k)(h(k) - h^2(m))$$

$$FNI_h(k,m,n_a) = (k-1)h(k) - (\eta_k - 1)h(m) + (n_a - m - (k-\eta_k))h(m)h(k) - (n_a - m)h^2(m)$$

$$J_h(m) = (m-1)[h(m+1) - h^2(m)]$$

THE STABILITY OF A 2-STAR ENVIRONMENT. — Proposition 4 implies that a 2-Star environment is OCS if the membership fee is high enough to preclude any subset of peripheral individuals from founding a new club that does not include the central individual

[14] If $\frac{m(m-1)}{n_a-1} < 2$ then $FNI_h(m,m,n_a) \geq J_h(m)$ and $J_h(m)$ is not the maximizing element of the lower bound.

and low enough that it is beneficial for the central individual to maintain each affiliation. Claim 2 summarizes the conditions for stability of a 2-Star environment for various characteristics of the club congestion function:

Claim 2. *Denote* $l_h = \min\{k \in \mathbb{Z} | h(k) \leq h^2(2)\}$.

(i) Let $h(\cdot)$ *be the club congestion function. The* 2*-Star environment is OCS if and only if* $h(2) \geq c \geq \max_{k \in \{2,\ldots,\min\{l_h - 1, n_a - 1\}\}} (k-1)(h(k) - h^2(2))$.

(ii) Let $h(\cdot)$ *be an elastic club congestion function. The* 2*-Star environment is OCS if and only if* $h(2) \geq c \geq h(2) - h^2(2)$.

(iii) Let $h(\cdot)$ *be the reciprocal club congestion function. The* 2*-Star environment is OCS if and only if* $c \in [0, 1]$.

(iv) Let $h(\cdot)$ *be the exponential club congestion function. The* 2*-Star environment is OCS if and only if* $a + \delta \geq c \geq \max_{k \in \{2,\ldots,\min\{l_h - 1, n_a - 1\}\}} (k-1)((a + \delta^{k-1}) - (a + \delta)^2)$. *If* $a = 0$*, the condition becomes* $c \in [\delta - \delta^2, \delta]$.

Individuals in a 2-Star environment suffer no congestion. Therefore, indirect paths in the 2-Star environment can only be improved by direct links. By forming a new club of size k that does not include the central individual, peripheral individuals only gain from direct links to other deviators, $(k-1)(h(k) - h^2(2))$. The second part shows that when the club congestion function is elastic there is always a range of membership fees wherein the 2-Star environment is OCS.[15] The third and fourth part characterize the existence of a membership fee wherein the 2-Star environment is OCS given specific club congestion functions.

m-STAR ENVIRONMENTS WITH $m > 2$. — m-Star environments where $m > 2$ are environments where individuals suffer both from the indirect connections depreciation friction and from the club congestion friction. Claim 3 provides two examples where such environment is Open Clubwise Stable.

Claim 3. *Let* $n_a \geq 9$.

(i) Let $h(\cdot)$ *be the reciprocal club congestion function. The* 3*-Star environment is OCS if and only if* $c = 1$.

(ii) Let $h(\cdot)$ *be the exponential club congestion function with* $a = 0$*. The* 3*-Star environment is OCS if and only if* $c \in [\delta + \delta^3 - 2\delta^4, 2\delta^2]$*. This range exists if and only if* $\delta \geq \frac{1}{2}$.

[15]Generally, the existence of such a range is not guaranteed. Consider, for example, the case where $h(2) = 0.3$, $h(3) = 0.25$ and $n_a \geq 4$. In this case, the central individual would abort her affiliations for every membership fee above 0.3. However, a triad of peripheral individuals will form a new club if the membership fee is lower than 0.32. Thus, if $h(2) = 0.3$, $h(3) = 0.25$ and $n_a \geq 4$, the 2-star environment is never OCS.

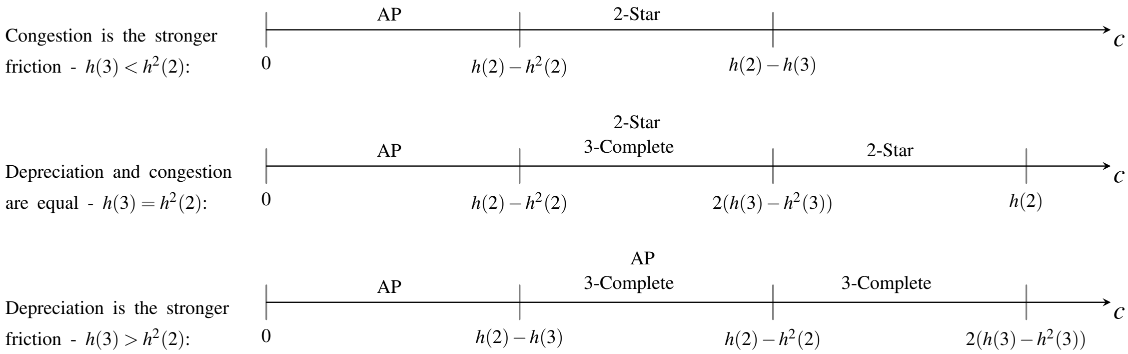


FIGURE 5. WEAK LINKS AS A FUNCTION OF THE MEMBERSHIP FEE.

When the club congestion function is exponential with $a = 0$, we prove that the most attractive new club is one formed by two peripheral individuals that do not share a club in the 3-Star environment. This club provides these individuals with a direct link between themselves and an improved indirect link to the non-central individual affiliated with their partner in the original environment. If the congestion friction is strong then the 3-Star environment is never OCS. On the one hand, strong congestion leads to small benefits accruing to the central individual from each affiliation. On the other hand, due to congestion the benefit of forming a new small club is relatively high. When congestion is weakened (δ increases) the incentive for the central individual to leave a club weakens since membership in the club becomes more profitable. In addition, peripheral individuals refrain from coalitional deviations since the links induced by the original clubs are satisfactory. Section 5 in the online appendix provides further analysis of the stability of m-Star environments.

E. The Emergence of Weak Links

As previously discussed (Section IV.C), as long as the membership fee is low enough, each individual is able to avoid both congestion and depreciation by forming intimate clubs with all other individuals. However, once the membership fee becomes higher, weak links emerge as low quality substitutes. There are two types of weak links in the club formation setup. One is based on cost-less indirect links and the other is based on larger clubs that induce low quality direct connections at low cost (per link) to substitute for costly intimate connections. Our analysis highlights the trade-off between these two types of weak links, that depends on the severity of the club congestion friction relative to the indirect links depreciation friction. This trade-off, therefore, can be captured by the relationship between $h(3)$ and $h^2(2)$: $h(3)$ represents the loss due to congestion in a club of size 3 while $h^2(2)$ stands for the loss of indirect connection that includes two links formed in clubs of size 2. The following discussion is summarized by Figure 5.

When club congestion is the stronger friction, that is $h(3) < h^2(2)$, the All Paired environment is OCS for low membership fee, but for a higher fee ($c \in (h(2) - h^2(2), h(2) - h(3))$) the 2-Star environment becomes OCS (while the 3-complete environment is not OCS). This result is in line with most of the strategic network formation literature. The frugal star architecture emerges as an equilibrium that efficiently maintains connectivity at much lower costs. By incorporating club formation into the setup of strategic network formation, we show that this prediction holds only if congestion is the predominant friction in the formation process.

Now we turn to the case where club congestion is the weaker friction, that is $h(3) > h^2(2)$. We assume that the congestion is not too strong, specifically, $h(3) \geq 0.15$.[16] Our previous results imply that in this case the 3-Complete environment, rather than the 2-Star environment emerges as OCS.[17] This result suggests a new insight - *larger clubs, that induce low quality direct connections at low cost (per link) turn out to be substitutes for costly intimate connections*. Hence, when congestion is less substantial than depreciation, weak ties are direct congested links. In addition, this implies that when depreciation is the stronger friction, complete networks may survive high maintenance costs by reducing the quality of the links.

When the two frictions are of the same magnitude, $h(3) = h^2(2)$, both 3-Complete and 2-Star environments are OCS once the membership fee is too high for the All Paired environment to be OCS. Note, however, that the 2-Star environment is a Minimally Connected environment while the 3-Complete environment is not. As a result, the marginal utility of each affiliation in the 2-Star environment is higher than that of the 3-Complete environment. Therefore, the range of costs wherein the 2-Star environment is OCS is larger than that wherein the 3-Complete environment is OCS (for $c \in (2(h(3) - h^2(3)), h(2))$ the 2-Star environment is OCS while the 3-Complete is not).

This case can be demonstrated by an exponential club congestion function with $a = 0$ (assuming δ is high enough). The All Paired environment is the unique OCS environment when the membership fee is very low and the 2-Star and the 3-Complete environments are OCS when c increases. But, when $\delta \geq c > 2(\delta^2 - \delta^4)$, while the 2-Star is OCS, the 3-Complete environment is not OCS since aborting existing affiliations becomes worthwhile as indirect connections are an attractive alternative.

Interestingly, even though congestion in the 3-Star environment is similar to that of the 3-Complete environment, leaving an existing club is not compensated by indirect connections. Indeed, by Claim 3, if $\delta \geq \frac{1}{2}$ then for slightly higher costs ($\delta + \delta^3 - 2\delta^4 > 2(\delta^2 - \delta^4)$), the 3-Star environment is OCS. In fact, when $c \in (\delta, 2\delta^2]$, the 2-Star environment is not OCS, while the 3-Star is OCS due to the higher value of a single affiliation to the central individual. This is another demonstration of the usefulness of weak ties

[16] The lower bound on $h(3)$ is necessary because for very small values of $h(3)$ the 3-Complete environment may not be OCS even if $h(3) \geq h^2(2)$. However, 0.15 is not a tight lower bound on $h(3)$. The exact condition is $h(3) > \max\{h^2(2), \frac{3}{4}[1 - \sqrt{1 - \frac{8h(2)}{9}}]\}$.

[17] The All Paired environment is OCS if and only if $c \in (0, h(2) - h^2(2))$ and it is not OCS when $c > h(2) - h^2(2)$. The 3-Complete environment is OCS if and only if $c \in [h(2) - h(3), 2[h(3) - h^2(3)]]$. If $c < h(2) - h^2(2)$ The 2-Star environment is not OCS. Using the properties of $f(x) = x - x^2$ it is possible to show that $h(3) > h^2(2)$ implies $2[h(3) - h^2(3)] > h(2) - h^2(2) > h(2) - h(3)$.

generated in big clubs.

Things are a bit different if analyzed from social welfare perspective. As was previously discussed, the All Paired environment is the SE environment for $c \in (0, \min\{h(2) - h^2(2), h(2) - h(3)\})$. When congestion is the stronger friction, once the membership fee becomes too high ($c > h(2) - h^2(2)$) the 2-Star environment is socially preferred to the All Paired and the 3-Complete environments. In fact, this is still the case when depreciation is the stronger friction ($h(3) > h^2(2)$) but congestion is still considerable (e.g. when $h(2) - h(3) > h(3) - h^2(2)$ and n_a is large). Thus, the 3-Complete environment achieve higher total utility than both the All Paired and the 2-Star environments for some range of membership fees only when depreciation is the stronger friction and congestion is relatively negligible.[18] Moreover, as the number of individuals grows, the 2-Star environment becomes more socially attractive compared to the 3-Complete environment since the number of clubs (and therefore membership fees) grows linearly in the former and quadratically in the latter. Hence, in some cases (e.g. n_a is large) individuals fail to internalize the effect of congestion and form the 3-Complete environment while the 2-Star is more desirable socially.

In fact, this gap between stability and efficiency is more general. Proposition 3 shows that the highest membership fee for which an m-Complete environment is OCS is $(m-1)(h(m) - h^2(m))$. Since strong efficiency implies satisfying the "No New Club Formation", Proposition 2 implies that there is never a case wherein an m-Complete environment is strongly efficient and not OCS. The opposite, however, as shown above, is possible. For m-Star environments recall that by Proposition 2, m-Star environments are efficient relative to m-Uniform environments for some range of membership fees such that $c = k_h(m)$ is always strictly included within this range. By Proposition 4, m-Star environments are never OCS when $c > k_h(m)$, meaning that a range of membership fees always exists where m-Star environments are not OCS although they are efficient relative to all m-Uniform environments (for numeric examples see Section 6 in the online appendix).

V. Club Rules: Closed Clubwise Stability

There are many possible rules regarding the forming, joining or leaving of social clubs. Each set of rules induces a different set of possible deviations and therefore corresponds to a different stability concept. So far we have only considered Open Clubwise Stability that implements an open environment in which joining, leaving and formation of clubs are done freely as long as the membership fee is paid. But there are environments in which clubs have more restrictive rules. For example, clubs in which the acceptance of new members requires the agreement of incumbent club members, are very com-

[18]The total utility in the All Paired environment is $n_a(n_a-1)(h(2)-c)$. The total utility in the 3-Complete environment is $\frac{n_a(n_a-1)}{2}(2h(3)-c)$. The total utility in the 2-Star environment is $(n_a-1)(h(2)-c)+(n_a-1)(h(2)+(n_a-2)h^2(2)-c)$. The All-paired environment dominates the 2-Star environment if and only if $c < h(2)-h^2(2)$ and the 3-Complete environment if and only if $c < 2(h(2)-h(3))$. The 3-Complete environment dominates the 2-Star environment if and only if $c < \frac{2n_a-4}{n_a-4}(h(3)-h^2(2)) - \frac{4}{n_a-4}(h(2)-h(3))$ that approaches $2(h(3)-h^2(2))$ from above when n_a grows larger.

mon in various social groups (e.g. academic departments, fraternities and sororities and kibbutzim) as well as in international organizations (e.g. the European Union). We demonstrate the application of this club rule by introducing the Closed Clubwise Stability solution concept which considers stability under the requirement that incumbents must unanimously approve every new member.[19]

An Environment G is ***Closed Clubwise Stable*** (henceforth, CCS) if the following conditions obtain:

(i) No Leaving: $\forall s \in S, \forall i \in N_G(s) : u_i(G,w,c) \geq u_i(G-\{i,s\},w,c)$.

(ii) No New Club Formation: $\forall m \subseteq N$:
$\exists i \in m : u_i(G+m,w,c) > u_i(G,w,c) \Rightarrow \exists j \in m : u_j(G+m,w,c) < u_j(G,w,c)$.

(iii) No Joining: $\forall s \in S, \forall i \notin N_G(s)$:
$u_i(G,w,c) \geq u_i(G+\{i,s\},w,c) \;\; OR \;\; \exists j \in N_G(s) : u_j(G,w,c) > u_j(G+\{i,s\},w,c)$.

For an environment to be CCS, the unanimous agreement of incumbent club members is required in order to join a club. Since this requirement makes the joining deviation harder to execute, Open Clubwise Stability is a refinement of Closed Clubwise Stability.

Generally, admission of new members into the club induces both positive and negative externalities upon incumbent members. Positive externalities stem from new (or shorter) paths provided by the new member. Negative externalities stem from the effects of congestion. Clearly when there is no congestion, incumbents receive only positive externalities from admitting new members, and therefore they would never object it. This implies that in the baseline case of no congestion the OCS and CCS solution concepts coincide. In addition, when there is no membership fee the set of CCS environments is clearly the set of spanning super environments of the All Paired environment. Hence, the difference between Open Clubwise Stability and Closed Clubwise Stability exists only when the membership fee is positive and club congestion exists.

In order to demonstrate the difference between OCS and CCS we define the **Almost Grand Club environment** as an environment in which there is exactly one populated club and all individuals except for one are affiliated with it. Let $n_a > 3$ and let $h(\cdot)$ be a club congestion function such that $k_h(n_a-1) > k_h(n_a) > \max_{k \in \{2,\ldots,n_a-2\}} k_h(k)$.[20] Consider the Almost Grand Club environment when $k_h(n_a) > c > \max_{k \in \{2,\ldots,n_a-2\}} k_h(k)$. In this case, no individual wants to leave the populated club since $k_h(n_a-1) > c$. No subset of indi-

[19]The literature on the stability of coalition partitions and jurisdictions also explores various admission rules. The basic rule is free mobility wherein each group of individuals can freely move from one coalition to another. Well studied restrictions of free mobility in this context are exclusion due to other affiliations, admission rules and capacity thresholds.

[20]One example of a club congestion function that satisfies these properties is the exponential congestion function with $a = 0$ and $\delta = \frac{\sqrt{4n_a^2-16n_a+14}}{2(n_a-1)}$.

viduals would want to form a new club since $c > \max_{k \in \{2,\ldots,n_a-2\}} k_h(k)$.[21] Since $k_h(n_a) > c$, the isolated individual wishes to join the populated club and the Almost Grand Club environment is not OCS. But, since $k_h(n_a-1) > k_h(n_a)$ such a deviation will strictly harm incumbents of the populated club who lose out because the positive externalities (one new direct connection) are lower than the negative externalities (weaker direct connections to all other incumbents due to stronger club congestion). Hence, the Almost Grand Club environment is CCS. Although we do not pursue a dynamic analysis of our setting wherein individuals join the environment sequentially, it is intuitive that while OCS encourages integration (in this example, one big club), CCS may drive the environment toward segregation (uniform partition).

VI. Real Life Implications: Homophily and Clustering

A. *Homophily*

When the population consists of different types of individuals, homophily becomes a major interest in the formation of social networks. In this literature, homophily is manifested by individuals establishing links only (or mostly) within their own type. In the traditional club theory the meaning of segregation (homophily) is that jurisdictions (clubs) are composed from homogeneous individuals.[22] The driver for homophily in both literatures is heterogeneity of preferences. Individuals of different types may have different preferences over local public goods or over the benefit from links to individuals of different types.[23]

Our setup of club formation is rich enough to generate homophily that is based on heterogeneous preferences. But, it may also give rise to homophily when the types have no preferences for discrimination (i.e. the value of a link is independent of the type of individuals it connects). To demonstrate this we consider a simple example in which the individuals are identical in their preferences with respect to the benefit from links, but they have different "social skills" which translate into different congestion functions.

Suppose that there are six individuals, four individuals are of type X and two individuals are of type Y. The two types differ only with respect to the congestion function. Type X has a constant congestion function $h_x(m) = \frac{1}{3}$ for every club size while type Y has a congestion function such that $h_y(2) = 1$ and $h_y(m) = 0$ for larger clubs. Note that the congestion function is defined on the size of club rather than on its composition. Consider an environment with two populated clubs A and B where the four type X individuals are members of club A and the two type Y individuals are members of club B. It turns

[21] For every size k of the new club, the benefits for individuals from the populated club are bounded from above by $k_h(k)$ (at least one of them should be a member of the new club). Therefore $c > \max_{k \in \{2,\ldots,n_a-2\}} k_h(k)$ guarantees no deviations to clubs of size $n_a - 2$ or smaller. The maximal gain for an individual affiliated with the populated club from being involved in the formation of a new club of size $n_a - 1$ or n_a is bounded by $h(n_a - 1)$. Since $n_a > 3$, $c > k_h(n_a - 2) = (n_a - 3)h(n_a - 2) \geq h(n_a - 1)$. This means that $c > \max_{k \in \{2,\ldots,n_a-2\}} k_h(k)$ guarantees no deviations to form new clubs.

[22] See Tiebout (1956) and Wooders (1980, 1989).

[23] Homophily may also be an outcome of different costs of association (e.g. Jackson and Rogers (2005)).

out that this environment is Open Clubwise Stable for $c \in [\frac{2}{3}, 1]$. The difference in the congestion functions is important for the (no) formation of a new club of size two that includes one type X individual and one type Y individual. The type Y individual that does not suffer congestion in clubs of size 2, would have been happy to form such a club due to the new direct and indirect links it implies. However, the type X individual suffers from congestion that makes it not worthwhile to form a new club while receiving only two connections (one direct and one indirect) in return.[24]

It is interesting to note that homophily due to differences in congestion functions is consistent with the evidence described in McPherson and Smith-Lovin (1982). They describe a homophilous environment where men (type X) tend to belong to much larger clubs than do women (type Y). Gender differences in various aspects of social life can be attributed either to the larger number of direct contacts formed by men or to the higher quality of direct contacts cultivated by women.

B. Clustering

A well-known real-life phenomenon in social networks is that they are characterized by high clustering. That is, in most real-life networks the probability of two individuals who share a common neighbor to be connected is much higher than would be expected if connections had been formed randomly (see Goyal (2007) and Jackson (2008)). High clustering affects the spread of information and therefore access to jobs, ideas and other resources.

Social sciences literature (see Rivera, Soderstrom and Uzzi (2010) for a recent survey) frequently attributes high clustering in social networks to one of two explanations. One explanation is based on homophily (see McPherson, Smith-Lovin and Cook (2001)). The other explanation assumes individual preference for connections with individuals with whom a shared connection already exists. Termed "preference for transitivity," it can be based on various motives, such as reduced uncertainty, improved monitoring, conflict mitigation and minimization of opportunism (see Heider (1946), Cartwright and Harary (1956), Coleman (1988) and Hummon and Doreian (2003)).

A relatively recent body of literature attempts to provide econometric tools for estimating network formation models that incorporate homophily, preference for transitivity and state dependence in links.[25] A growing concern in this literature is that neglect of self-selection into social contexts leads to over-estimation of the importance of homophily and preference for transitivity in the process of network formation (see Rivera, Soderstrom and Uzzi (2010), Currarini, Jackson and Pin (2010) and Miyauchi (2016)).

Indeed, we believe that our setting provides a third explanation for the high cluster-

[24]Complete segregation, as in this example, is not a general feature in network formation models with heterogeneity since the gain from one link between two components guarantees connections (mostly indirect) to all members of the other component. Therefore, it usually very beneficial in these models to initiate few links across components.

[25]One of the main challenges of this literature is the treatment of homophily on unobservables. Goldsmith-Pinkham and Imbens (2013) introduce homophily on unobservables by assuming that the relevant unobservables are binary and distributed independently of all observables. In Mele (2018) individuals are partitioned exogenously to unobserved communities and they exhibit preference for transitivity only within these communities. Graham (2015, 2016) proposes to exploit the fact that homophily is independent of network structure. See the discussion in Jackson (2014).

ing observed in real-life networks. Since every pair of individuals who share a club is connected in the induced network, the affiliation portfolios chosen by individuals induce a social network composed of a collection of cliques. Therefore, in our framework, a network induced by non-trivial clubs (e.g. of size greater than 2) must exhibit high local clustering since an individual's neighbors form a tightly knit group (see also Jackson, Rodriguez-Barraquer and Tan (2012)). Hence, we propose to consider clubs as linking platforms rather than individuals' linking preferences as the fundamental that drives the high clustering observed in real-life networks.[26]

VII. Concluding comments

This paper focuses on the formation of social networks based on the endogenous formation of social clubs and in particular on the role of clubs as platforms for link formation. Most of our analysis relies on the assumptions that clubs (beside their size) and individuals are homogeneous. A more complete picture of the social architecture may include the endogenous formation of a variety of clubs that may differ in membership costs, quality of induced links and rules of entry, exit and formation. For example, in some clubs the interaction among members may be more intense than in others and as a result they may differ in their congestion functions and membership fees. We also assume that individuals are homogeneous. As we demonstrate by example in Section VI.A, in a model of heterogeneous individuals the weight of each link may depend on the identity of the individuals and may be asymmetric. In addition, individuals may exhibit preferences for discrimination and attention capacity constraints. These costs and benefits affect the attractiveness of the different clubs and their composition. As a result they also affect the stable social environments that may emerge from our endogenous affiliation setting.

REFERENCES

Arnold, Tone, and Myrna Wooders. 2015. "Dynamic club formation with coordination." *Journal of Dynamics and Games*, 2(3&4): 341–361.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. "Who's Who in Networks. Wanted: The Key Player." *Econometrica*, 74(5): 1403–1417.

Bar, Talia. 2005. "Coalition Networks and Induced Graphs." Working Paper.

[26]Obviously, in some contexts these explanations coexist. One example is the discussion in Currarini, Jackson and Pin (2009) on social clubs as the platform on which matching biases (as homophily) evolve. Another example is Kossinets and Watts (2006) who track emails of students, faculty, and staff at a large research university over an academic year. They find that among students who do not share a common class, having a mutual contact increase the probability of communication by 140 times. However, if students do share a class they were only 3 times more likely to begin communicating if they shared a common correspondent. Datasets that contain both club affiliations and the social network, as in the one studied by Kossinets and Watts (2006) (or by Young and Larson (1965*a*)), may enable researchers to disentangle these three explanations (see also the discussion in Feld (1981)). The method suggested in Chandrasekhar and Jackson (2018) for the estimation of network formation models, can be interpreted as an econometric analysis of network data where club affiliations are assumed but not observed (see Section 3 therein).

Berge, Claude. 1989. *Hypergraphs: Combinatorics of Finite Sets.* Vol. 45 of *North-Holland Mathematical Library*, Elsevier Science Publishers B. V.

Bloch, Francis, and Bhaskar Dutta. 2009. "Communication Networks with Endogenous Link Strength." *Games and Economic Behavior*, 66(1): 39–56.

Borgs, Christian, Jennifer Chayes, Jian Ding, and Brendan Lucier. 2011. "The hitchhiker's guide to affiliation networks: A game-theoretic approach." In *Innovations in Computer Science - ICS 2011, Tsinghua University, Beijing, China.* , ed. Bernard Chazelle, 389–400. Tsinghua University Press.

Brueckner, Jan K. 2006. "Friendship Networks." *Journal of Regional Science*, 46(5): 847–865.

Buchanan, James M. 1965. "An Economic Theory of Clubs." *Economica, New Series*, 32(125): 1–14.

Cabrales, Antonio, Antoni Calvó-Armengol, and Yves Zenou. 2011. "Social interactions and spillovers." *Games and Economic Behavior*, 72(2): 339–360.

Calvó-Armengol, Antoni. 2004. "Job contact networks." *Journal of Economic Theory*, 115(1): 191–206.

Calvó-Armengol, Antoni, and Yves Zenou. 2003. "Does crime affect unemployment? The role of social networks." *Annales d'Economie et de Statistique*, 71–72: 173–188.

Cartwright, Dorwin, and Frank Harary. 1956. "Structural Balance: A Generalization of Heider's Theory." *Psychological Review*, 63(5): 277.

Caulier, Jean-François, Ana Mauleon, and Vincent Vannetelbosch. 2015. "Allocation rules for coalitional network games." *Mathematical Social Sciences*, 78: 80–88.

Caulier, Jean-François, Ana Mauleon, José J. Sempere-Monerris, and Vincent Vannetelbosch. 2013. "Stable and Efficient Coalitional Networks." *Review of Economic Design*, 17: 249–271.

Chandrasekhar, A., and M. O. Jackson. 2018. "A Network Formation Model Based on Subgraphs." Working Paper.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology*, 94: S95–S120.

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

Currarini, Sergio, Matthew O Jackson, and Paolo Pin. 2010. "Identifying the roles of race-based choice and chance in high school friendship network formation." *Proceedings of the National Academy of Sciences*, 107(11): 4857–4861.

Faust, Katherine. 1997. "Centrality in affiliation networks." *Social networks*, 19(2): 157–191.

Feld, Scott L. 1981. "The focused organization of social ties." *American Journal of Sociology*, 86(5): 1015–1035.

Fershtman, Chaim, and Neil Gandal. 2011. "Direct and Indirect Knowledge Spillovers: the 'Social Network' of Open-Source Projects." *RAND Journal of Economics*, 42(1): 70–91.

Goldsmith-Pinkham, Paul, and Guido W Imbens. 2013. "Social networks and the identification of peer effects." *Journal of Business & Economic Statistics*, 31(3): 253–264.

Goyal, Sanjeev. 2005. "Strong and Weak Links." *Journal of the European Economic Association*, 3(2-3): 608–616.

Goyal, Sanjeev. 2007. *Connections: An Introduction to the Economics of Networks.* Princeton University Press.

Goyal, Sanjeev, Aleander Konovalov, and José Luis Moraga-González. 2008. "Hybrid R&D." *Journal of the European Economic Association*, 6(6): 1309–1338.

Graham, Bryan S. 2015. "Methods of identification in social networks." *Annual Review of Economics*, 7(1): 465–485.

Graham, Bryan S. 2016. "Homophily and Transitivity in Dynamic Network Formation." *Working Paper*.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *The American Journal of Sociology*, 78(6): 1360–1380.

Granovetter, Mark. 1983. "The Strength of Weak Ties: A Network Theory Revisited." *Sociological Theory*, 1: 201–233.

Haller, Hans. 2016. "Networks as public infrastructure: Externalities, efficiency, and implementation." *Journal of Public Economic Theory*, 18(2): 193–211.

Heider, Fritz. 1946. "Attitudes and cognitive organization." *The Journal of Psychology*, 21(1): 107–112.

Hummon, Norman P, and Patrick Doreian. 2003. "Some Dynamics of Social Balance Processes: Bringing Heider Back into Balance Theory." *Social Networks*, 25(1): 17–49.

Jackson, Matthew O. 2008. *Social and Economic Networks.* Princeton University Press.

Jackson, Matthew O. 2014. "Networks in the understanding of economic behaviors." *The Journal of Economic Perspectives*, 28(4): 3–22.

Jackson, Matthew O, and Asher Wolinsky. 1996. "A Strategic Model of Social and Economic Networks." *Journal of Economic Theory*, 71(1): 44–74.

Jackson, Matthew O., and Brian W. Rogers. 2005. "The Economics of Small Worlds." *Journal of the European Economic Association*, 3(2-3): 617–627.

Jackson, Matthew O, Tomas Rodriguez-Barraquer, and Xu Tan. 2012. "Social capital and social quilts: Network patterns of favor exchange." *American Economic Review*, 102(5): 1857–97.

Jun, Tackseung, and Jeong-Yoo Kim. 2009. "Hypergraph formation game." *Hitotsubashi Journal of Economics*, 50(2): 107–122.

Kadushin, Charles. 1966. "The Friends and Supporters of Psychotherapy: on Social Circles in Urban Life." *American Sociological Review*, 31(6): 786–802.

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

Kramarz, Francis, and Oskar Nordström Skans. 2014. "When strong ties are strong: Networks and youth labour market entry." *Review of Economic Studies*, 81(3): 1164–1200.

McGuire, Martin. 1974. "Group segregation and optimal jurisdictions." *Journal of Political Economy*, 82(1): 112–132.

McPherson, J Miller, and Lynn Smith-Lovin. 1982. "Women and weak ties: Differences by sex in the size of voluntary organizations." *American Journal of Sociology*, 87(4): 883–904.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology*, 27(1): 415–444.

Mele, Angelo. 2018. "A Structural Model of Homophily and Clustering in Social Networks." *Working Paper*.

Miyauchi, Yuhei. 2016. "Structural estimation of pairwise stable networks with nonnegative externality." *Journal of Econometrics*, 195(2): 224–235.

Montgomery, James D. 1992. "Job search and network composition: Implications of the strength-of-weak-ties hypothesis." *American Sociological Review*, 57(5): 586–596.

Montgomery, James D. 1994. "Weak ties, employment, and inequality: An equilibrium analysis." *American Journal of Sociology*, 99(5): 1212–1236.

Myerson, Roger B. 1980. "Conference Structures and Fair Allocation Rules." *International Journal of Game Theory*, 9(3): 169–182.

Page, Frank H., and Myrna Wooders. 2007. "Networks and Clubs." *Journal of Economic Behavior and Organization*, 64: 406–425.

Rivera, Mark T, Sara B Soderstrom, and Brian Uzzi. 2010. "Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms." *Annual Review of Sociology*, 36: 91–115.

Simmel, Georg. 1908/1955. "The Web of Group-Affiliations." In *Conflict and the Web of Group-Affiliations*. The Free Press. (Translated by Reinhard Bendix in 1955. Original appeared in "Soziologie" published by Duncker and Humblot, Berlin, 1908).

Slikker, Marco, and Anne Van den Nouweland. 2001. *Social and Economic Networks in Cooperative Game Theory.* Kluwer Academic Publishers.

Snijders, Tom AB, Philippa E Pattison, Garry L Robins, and Mark S Handcock. 2006. "New specifications for exponential random graph models." *Sociological methodology*, 36(1): 99–153.

So, Chiu Ki, Vai-lam Mui, and Birendra Rai. 2017. "Event Subscription and Noncooperative Network Formation." *Working paper*.

Tiebout, Charles M. 1956. "A Pure Thoery of Local Expenditures." *The Journal of Political Economy*, 64(5): 416–424.

Wooders, Myrna. 1978. "Equilibria, the core, and jurisdiction structures in economies with a local public good." *Journal of Economic Theory*, 18(2): 328–348.

Wooders, Myrna. 1980. "The Tiebout Hypothesis: Near Optimality in Local Public Good Economies." *Econometrica*, 48(6): 1467–1485.

Wooders, Myrna Holtz. 1988. "Stability of jurisdiction structures in economies with local public goods." *Mathematical Social Sciences*, 15(1): 29–49.

Wooders, Myrna Holtz. 1989. "A Tiebout Theorem." *Mathematical Social Sciences*, 18(1): 33–55.

Young, Ruth C, and Olaf F Larson. 1965*a*. "The contribution of voluntary organizations to community structure." *American Journal of Sociology*, 71(2): 178–186.

Young, Ruth C., and Olaf F. Larson. 1965*b*. "A New Approach to Community Structure." *American Sociological Review*, 30(6): 926–934.