A Micro-foundational Model and Analysis of Evolving Social Networks

Ahmed M. Alaa, Kartik Ahuja, and Mihaela van der Schaar
Department of Electrical Engineering, UCLA
Emails: {ahmedmalaa@ucla.edu, ahujak@ucla.edu,mihaela@ee.ucla.edu}

Abstract—Many societies are organized in networks. Real-world social networks such as friendship networks, online social networks, scientific collaboration and citation networks, are formed by people who meet and interact over time. The way people meet is highly influenced by the evolving network structure, and their decisions to connect depend mainly on their intrinsic characteristics. In this paper, we present a first mathematical model to capture the microfoundations of social networks evolution, where people modeled as boundedly rational agents of different types join the network, meet other agents stochastically over time, and consequently decide to form a set of *social ties*. Based on the meeting process, which is governed by the level of *structural opportunism* and the a priori *type distribution*, as well as the incentives of the agents to form links, which are governed by *homophily* and *social gregariousness*, agents make a sequence of link formation decisions that lead to an endogenously evolving network. A basic premise of our model is that in real-world networks, agents do not form links in a one-shot fashion via a *preferential attachment* probabilistic rule, but rather form links by reasoning about the social benefits that agents they meet over time can bestow. We analytically study the evolution of the endogenously formed networks in terms of friendship and popularity acquisition given the following exogenous parameters: structural opportunism, type distribution, homophily, and social gregariousness. We show that the time needed for an agent to find "friends" is highly influenced by the exogenous network parameters: agents who are more gregarious, more homophilic, less opportunistic, or belong to a type "minority" spend, on average, a longer time searching for friends. Moreover, we show that preferential attachment and thus, an agent's popularity acquisition, is a direct consequence of an endogenously emerging *preferential meeting* process, in which agents who search for friendships meet more popular agents with higher probability. We also show that the meeting process can be *doubly preferential*, in which agents of a certain type meet more popular similar-type agents with higher probability. Such meeting process creates asymmetries in the levels of popularity attained by different types of agents.

Fig. 1: Framework for the analysis of social networks evolution.

I. INTRODUCTION

We are living in the era of networks. With the widespread usage of online social networking (OSN) platforms, such as Facebook and Twitter [1]; academic networking websites such as ResearchGate [2]; and professional online networks such as Linkedin [3], people are getting more and more connected. With people interacting and getting connected through these platforms, networks emerge endogenously as a result of the actions of people who meet others over time, and take link formation decisions, i.e. "follow" a user on Twitter, "add" a friend on Facebook, "cite" a paper that is indexed by 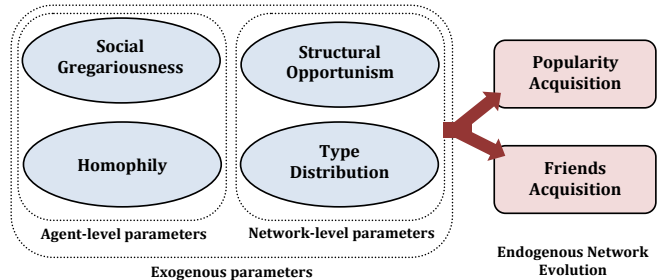Google Scholar, or "collaborate" with a researcher. Examples of emerging networks include: friendship networks [4] [5] [6], scientific collaboration and citation networks [7], and professional networks [3]. Understanding how networks form and evolve is essential for drawing insights into networked social interactions, carrying out predictions, and designing policies that can guide network formation. While extensive research has been recently devoted to the study of social networks, no systematic model exists that can explain how networks form and evolve over time based on the individuals decisions and preferences.

In this paper, we present a comprehensive micro-foundational model and analysis for dynamic social network formation. In our model, networks are formed over time by the actions of boundedly rational agents that arrive at the network stochastically, and meet other agents via a random process that is itself highly influenced by the dynamic network structure and the characteristics of the agents themselves. Thus, networks evolve over time as a stochastic process driven by the individual agents, where the formation of *social ties* among agents are in part endogenously determined, as a function of the current network structure itself, and in part exogenously, as a function of the individual characteristics of the agents. Agents have bounded rationality, i.e. they only have information about other agents they meet over time, and they are not able to observe the global network structure or reason about links formed by others. We focus on the impact of various exogenous parameters that describe both the characteristics of individuals forming the network, and the nature of the network itself, on the endogenously evolving network structure. While many network metrics such as diameter, clustering coefficient [14], community structures [7], and degree distribution [15] can be computed

using our model, we focus on two basic aspects of network evolution that describe the agent-level experience in the network: *friendship acquisition* (the process of forming links) and *popularity acquisition* (the process of gaining links), and we show how these experiences depend on the parameters considered. Before presenting our model and results, we provide the following definitions for the exogenous parameters under study. Fig. 1 depicts and categorizes these parameters.

1- Type Distribution: Agents are heterogenous in the sense that they possess type attributes that correspond to their preferences, race, ethnicity, etc. The experiences of different types of agents in the network are generally not symmetric. The type distribution characterizes the fraction of agents of each type in the network. We say that an agent belongs to a *type minority* to qualitatively describe a scenario where the fraction of agents of the corresponding type in the population is small, and we say that an agent belongs to a *type majority* otherwise.

2- Homophily: A pervasive feature of social networks that corresponds to the tendency of the agents to connect to other similar-type agents [9] [10] [11]. The extent to which a certain type of agents is *homophilic* is captured by an *exogenous homophily index*, which we formally define in section II. The homophily index can be thought of as the amount of "intolerance" that a certain type of agents have towards making contacts with other types.

3- Social Gregariousness: Some types of agents can be more *social* than others, and thus are willing to form more links. Social gregariousness is simply measured by the minimum number of links an agent is willing to make.

4- Structural Opportunism: Agents in the network are said to be *opportunistic* if they exploit their contacts to find new contacts; thus, agents are more likely to link with the friends of their friends if they are opportunistic. Structural opportunism can also be interpreted as a *social norm* that agents are expected to follow. For instance, users in Twitter are expected to *retweet* the tweets posted by users they follow, which leads to the followers of followers of a certain user to follow him. Structural opportunism can also be a social norm in friendship networks, where people introduce their friends to each other or people enjoy/trust the friends of their friends more than strangers.

A. Preview of the results

The central finding of this paper is that an agent's experience in the network is affected by the agents it meets over time. This meeting process is in turn affected by the exogenous network parameters. We classify our results as follows:

1- Friendship acquisition characterization: Agents joining the network will form a finite number of links over time. We say that agent A is a *friend* of agent B if agent B forms a directed link with agent A. In section III-A, we study the impact of the exogenous parameters on the time needed for an agent to find its *friends*. We show that agents who are more gregarious, more homophilic, less opportunistic, or belong to a

type "minority" are more likely to spend more time searching for friendships.

2- Popularity acquisition characterization: After an agent joins the network, other agents form links with it. The number of such links reflects the level of popularity of that agent, e.g. number of followers in Twitter, and number of citations in a citation network. We say that agent A is a *follower* of agent B if agent A forms a directed link with agent B. In section III-B, we study the impact of the exogenous parameters on the popularity acquisition time and popularity growth rates in a large network. We show that popularity evolution depends on the meeting process, which can in general be *doubly preferential*, i.e. the meeting process is both type and popularity biased. We prove that depending on the exogenous parameters, preferential attachment, which corresponds to the cumulative advantage in popularity acquisition, emerges as a result of the preferential meeting process. Moreover, we show that in homophilic societies, an emerging doubly preferential meeting process allows more gregarious agent types to get more popular, while in non-homophilic societies, an agent's age in the network determines its popularity over time.

B. Related works

Previous works on network formation can be divided into three categories: networks formed based on *random events* [8], [11]-[21], networks formed based on *strategic decisions* [22]-[26], and empirical models distilled by mining networks' data [4]-[7], [30]. A fairly large body of literature has been devoted to developing mathematical models for network formation, yet much fewer works attempt to interpret and understand how networks evolve over time and how can individual agents affect the characteristics of such networks. Probabilistic models based on random events are generative models that are concerned with constructing networks that mimic real-world social networks. In [11]-[18], agents get connected in a pure probabilistic manner in order to realize some degree distribution [12], or according to a *preferential attachment* rule [13] [14]. While such models can capture the basic structural properties of social networks, they fail to explain why and how such properties emerge over time. In contrast, strategic network formation models such as those in [22]-[26], and our previous works in [27] [28], can offer an explanation for why certain network topologies emerge as an equilibrium of a network formation game. However, these results are limited to studying network *stability* and *efficiency*, and provide only very limited insight into the dynamics and evolution of networks. Finally, mining empirical data can help building algorithms for detecting communities [31]-[34], predicting agents' popularity [30], or identifying agents in a network [29], but cannot help understanding how networks form and evolve.

II. MODEL

A. Network model

We consider a discrete-time model for a growing social network where one *agent* is born each time step and is indexed

by its birth date $i \in \{1, 2, \ldots, t, \ldots\}$. At date $t \in \mathbb{N}$, a snapshot of the network is modeled by a *step graph* $\mathcal{G}^t$ given by $\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)$, where $\mathcal{V}^t$ is the set of nodes, $\mathcal{E}^t = \{e_1^t, e_2^t, \ldots, e_{|\mathcal{E}^t|}^t\}$ is the set of edges between different nodes, with each edge e_k^t being an ordered pair of nodes $e_k^t = (i, j)$ ($i \neq j$, and $i, j \in \mathcal{V}^t$), and $|\mathcal{E}^t|$ is the number of distinct edges in the graph. Thus, $\mathcal{G}^t$ is a directed graph. Nodes correspond to agents (social actors) and edges correspond to directed links (social ties) between the agents. The adjacency matrix of $\mathcal{G}^t$ is denoted by $\mathbf{A}_\mathcal{G}^t = [A^t(i, j)]$, $A^t(i, j) \in \{0, 1\}$, $A^t(i, i) = 0, \forall i, j \in \mathcal{V}^t$. An entry of the adjacency matrix $A^t(i, j) = 1$ if $(i, j) \in \mathcal{E}_k^t$, and $A^t(i, j) = 0$ otherwise. If $A^t(i, j) = 1$, then agent i initiates a link with agent j, and we say that j is a "friend" of i, and i is a "follower" of j. The directed nature of a link indicates the agent initiating the link, and only this agent obtains the *social benefit* of linking and pays the link cost. The *indegree* of agent i is the number of links that are initiated towards i, denoted by $\deg_i^-(t)$, while the *outdegree*, denoted by $\deg_i^+(t)$, is the number of links initiated by agent i.

Each agent $i \in \mathcal{V}^t$ in the network possesses a type attribute θ_i, which belongs to a finite set of types $\theta_i \in \Theta, \Theta = \{1, 2, 3, \ldots, |\Theta|\}$, where $|\Theta|$ is the number of types. The set of type-k agents at time t is denoted by $\mathcal{V}_k^t$, where $\mathcal{V}^t = \bigcup_{k=1}^{|\Theta|} \mathcal{V}_k^t$. Agents are identified by both their birth dates and types. The network starts initially with a seed graph $\mathcal{G}^0$, which we assume to be an empty graph with no agents at $t = 0$, and agents arrive one at a time at each date t. The agents arrival in the network is modelled as a stationary stochastic process $\lambda(t) = \{\theta_t\}_{t \in \mathbb{N}}$, with a sample space $\Lambda = \Theta^\mathbb{N}$, i.e. $\Lambda = \{(\theta_1, \theta_2, \ldots) : \theta_t \in \Theta, \forall t \in \mathbb{N}\}$. We assume that the types of agents are independent and identically distributed, and that the agents' *type distribution* is $\mathbb{P}(\theta_i = k) = p_k$, where $\sum_{k \in \Theta} p_k = 1$. Thus, $\lambda(t)$ is a *Bernoulli scheme*. Using *Borel's law of large numbers*, we know that

$$\mathbb{P}\left(\lim_{t \to \infty} \frac{1}{t} |\mathcal{V}_k^t| = p_k\right) = 1.$$

In other words, for a sufficiently large network size (and age t), the actual fraction of agents of each type in the network converges almost surely to the prior type distribution of the Bernoulli scheme. Thus, at date t, and for a large enough network, the expected number of type k agents in the network is $p_k t$, and the total number of agents is t, i.e. $|\mathcal{V}^t| = t$, $\mathbb{E}\{|\mathcal{V}_k^t|\} = p_k t$, and $\lim_{t \to \infty} \frac{|\mathcal{V}_k^t|}{|\mathcal{V}^t|} = p_k$.

B. The meeting process

Let $M_i(t) = \{m_i(t)\}_{t=i}^{i+T_i-1}$ be the *meeting process* of agent i, which corresponds to the sequence of birth dates of the agents that agent i meets over time (note that birth dates are used to identify agents), and T_i is the stopping time of $M_i(t)$, which we define in section III-A as the *link formation time*. The sample space of the meeting process is given by $\mathcal{M} = \{(m_i(i), m_i(i+1), \ldots, m_i(i+T_i-1)) : m_i(t) \leq t, m_i(t) \neq i\}$. Unlike the arrival process, which is stationary and exogenous, the meeting process depends on the history of actions of all

agents in the network, i.e. the probability that agent i meets agent j at time t depends on their relative positions in the network at time t, which in turn depend on the sequence of meetings for both agents up to time $t - 1$. Moreover, the probability that a certain sample path of the meeting process occurs depends on all the exogenous parameters shown in Fig. 1. We denominate the set of agents to whom agent i forms links as agent i's *friends*, and the set of agents that form links with agent i by agent i's *followers*. Denote the set of type-k friends of agent $i \in \mathcal{V}^t$ by $\mathcal{N}_{i,t}^{+,k}$, and the set of all friends of i as $\mathcal{N}_{i,t}^+ = \bigcup_{k=1}^{|\Theta|} \mathcal{N}_{i,t}^{+,k}$, where $|\mathcal{N}_{i,t}^+| = \deg_i^+(t)$. Similarly, we denote the followers of agent i by $\mathcal{N}_{i,t}^-$, where $|\mathcal{N}_{i,t}^-| = \deg_i^-(t)$. Define the set $\mathcal{K}_{i,t} = \bigcup_{j \in \mathcal{N}_{i,t-1}^+} \mathcal{N}_{j,t-1}^+ / \{i\}$ as the set of *friends of friends* of agent i at time t, and the set $\bar{\mathcal{K}}_{i,t} = \mathcal{V}^t / \{\mathcal{K}_{i,t} \bigcup \mathcal{N}_{i,t}^+ \bigcup i\}$ as the set of *strangers* to agent i at time t. We capture the degree of structural opportunism of agents in the society by a parameter[1] $\gamma \in [0, 1]$, where $\gamma = 0$ corresponds to fully opportunistic agents, and $\gamma = 1$ corresponds to non-opportunistic agents. That is, γ is a measures of how often an agent i finds new friends without exploiting its current connections as we show in the following meeting process. For $t \geq i$, agent i meets *one* agent selected uniformly at random from the set of friends of friends with probability $1 - \gamma$ if $\mathcal{K}_{i,t} \neq \phi$, while if $\bar{\mathcal{K}}_{i,t} = \phi$, then agent i meets one agent selected from $\mathcal{K}_{i,t}$ with probability 1, i.e.

$$\mathbb{P}\left(m_i(t) \in \mathcal{K}_{i,t} \mid \mathcal{K}_{i,t} \neq \phi, \bar{\mathcal{K}}_{i,t} \neq \phi\right) = 1 - \gamma,$$

$$\mathbb{P}\left(m_i(t) \in \mathcal{K}_{i,t} \mid \mathcal{K}_{i,t} \neq \phi, \bar{\mathcal{K}}_{i,t} = \phi\right) = 1.$$

On the other hand, agent i meets one agent selected uniformly at random from the set of strangers with probability γ if $\bar{\mathcal{K}}_{i,t} \neq \phi$, while if $\mathcal{K}_{i,t} = \phi$, then agent i meets one agent selected from $\bar{\mathcal{K}}_{i,t}$ with probability 1, i.e.

$$\mathbb{P}\left(m_i(t) \in \bar{\mathcal{K}}_{i,t} \mid \mathcal{K}_{i,t} \neq \phi, \bar{\mathcal{K}}_{i,t} \neq \phi\right) = \gamma,$$

$$\mathbb{P}\left(m_i(t) \in \bar{\mathcal{K}}_{i,t} \mid \mathcal{K}_{i,t} = \phi, \bar{\mathcal{K}}_{i,t} \neq \phi\right) = 1.$$

Moreover, other agents can occasionally meet agent i at each time step, i.e. $\mathbb{P}\left(m_j(t) = i \mid j \in \mathcal{V}^t / \{i\}, t \leq j + T_j - 1\right) > 0, \forall t \geq i$.

C. Agents' actions and utility functions

When agent i meets agent $m_i(t)$ at time t, it observes its type and decides whether or not to form a link with that agent. Agents draw social benefits by connecting to others, but those benefits are type-dependent and link formation is costly. Links are formed unilaterally and only the agent initiating the link bears a cost of c and attains the linking benefit. We assume *local externalities*, i.e. linking benefits do not flow to indirect contacts. Let the benefit attained by agent i from linking to agent j be $\alpha_{\theta_i \theta_j}$, where $\alpha_{\theta_i \theta_j} \in \mathbb{R}^+$. The action of node i at time t is $a_i^t \in \{0, 1\}$, where $a_i^t = 1$ indicates that node i

---

[1]We assume that all types of agents have the same γ. The analysis can be easily extended to the case when each type has a different γ.

forms a link with node $m_i(t)$. The utility function of agent i at date t is given by

$$u_i^t\left(\mathbf{a}_i^t\right) = v_i\left(\sum_{j \in \mathcal{N}_{i,t}^+} \alpha_{\theta_i \theta_j}\right) - c\sum_{k=i}^t a_i^k, \qquad (1)$$

where $\mathbf{a}_i^t = \left(a_i^i, a_i^{i+1}, \ldots, a_i^t\right)$, and $v(x) : x \to \mathbb{R}^+$ is the benefit function that measures the *social capital* [2]. We assume that $v(x)$ is concave[3], twice continuously differentiable, and monotonically increasing in x, and $v(0) = 0$. That is, the marginal benefit of forming links diminishes as the number of links increases. This corresponds to the fact that agents do not form an infinite number of links in the network, but rather form a "satisfactory" number of links [4]. The action taken by agent i depends on the marginal benefit of forming a link and the linking cost as shown in (2), where agent i forms a link to the agent it meets only if the marginal utility from linking is positive. Thus, agents are *myopic* and form links without taking future meetings into account. To capture the impact of *homophily*, we assume that $\alpha_{\theta_i \theta_k} > \alpha_{\theta_i \theta_j}, \forall\, |\theta_k - \theta_i| < |\theta_j - \theta_i|$, and $\alpha_{\theta_i \theta_k} = \alpha_{\theta_i \theta_j}, \forall\, |\theta_k - \theta_i| = |\theta_j - \theta_i|$.

D. The exogenous parameters

In order to measure the exogenous homophilic tendency of a certain type of agents, we propose a novel definition of an *exogenous homophily index* for type-k agents h_k, which is a variant of the well known *Coleman homophily index* [21]:

$$h_k \triangleq 1 - \frac{\sum_{m \neq k}^{|\Theta|} p_m \lim_{t \to \infty} r_{\theta_i}(m,t)}{1 - p_k}, \forall \theta_i = k \qquad (3)$$

where $r_{\theta_i}(m,t)$ is the *maximum excess representation* of type-m agents in agent i's friends at time t, which is given by

$$r_{\theta_i}(m,t) \triangleq \sup_{\mathcal{N}_{i,t}^+} \frac{|\mathcal{N}_{i,t}^{+,m}|}{\deg_i^+(t)}, \qquad (4)$$

where $r_{\theta_i}(k,t) = 1, \forall \theta_i = k$, and $0 \le h_k \le 1, \forall k \in \Theta$. When type-$k$ agents are indifferent to the types of agents it connects to, i.e. type-k agents are extremely non-homophilic, then we have $\lim_{t \to \infty} r_{\theta_i}(m,t) = 1, \forall \theta_i = k, m \in \Theta$, which means that $h_k = 0$. On the other hand, if agents restrict their links to same-type agents only, then $\lim_{t \to \infty} r_{\theta_i}(m,t) = \delta(m,k), \forall \theta_i = k, m \in \Theta$, where $\delta(x,y)$ is the *Kronecker delta* function, which means that $h_k = 1$. Now we compute the exogenous homophily index in closed-form. Let $L_{\theta_i}^*(k,\alpha) \in \mathbb{Z}$ be the maximum number of links with type-k agents that agent i can form in the time period $[T,\infty)$ given that $\sum_{j \in \mathcal{N}_{i,T-1}^+} \alpha_{\theta_i \theta_j} = \alpha$, i.e. $L_{\theta_i}^*(k,\alpha) \triangleq \sup\left(\lim_{t \to \infty}\left|\mathcal{N}_{i,t}^{+,k}\right| - \left|\mathcal{N}_{i,T-1}^{+,k}\right|\right)$, for $\sum_{j \in \mathcal{N}_{i,T-1}^+} \alpha_{\theta_i \theta_j} =$

α. This can be easily computed in closed-form by taking the first derivative of the inverse of the concave benefit function in (1). It can be shown that

$$\inf_{M_i(t) \in \mathcal{M}} \lim_{t \to \infty} \deg_i^+(t) = L_{\theta_i}^*(\theta_i, 0),$$

and the exogenous homophily index of agent i is given by[5]

$$h_{\theta_i} = \frac{1}{1 - p_{\theta_i}}\left(1 - \sum_{k=1}^{|\Theta|} p_k \frac{L_{\theta_i}^*(k,0)}{L_{\theta_i}^*(k,0) + L_{\theta_i}^*(\theta_i, \alpha_{\theta_i k} L_{\theta_i}^*(k,0))}\right).$$

The parameter $L_{\theta_i}^*(\theta_i, 0)$ represents the minimum number of links an agent can form; this parameter captures social gregariousness. In summary, our model captures the four exogenous parameters defined in section I as follows:

- *Homophily*: the homophily of type-k agents is captured by the exogenous homophily index h_k.
- *Social gregariousness*: the gregariousness of type-k agents is captured by $L_k^*(k,0)$.
- *Structural opportunism*: the parameter γ reflects the extent of structural opportunism.
- *Type distribution*: the fraction of type-k agents in a large network is given by p_k.

In the next section, we study the friendships acquisition experience for agents in the network.

III. Friendships acquisition: how long does it take to find friends?

In the next subsection, we characterize the time needed for agents to find their friends in the evolving network. Unlike previous works where link formation is a one-shot process (which is the case in [8], [14], [18], [19], [20], [21], [22], [23], [25], and [26]), in our model agents form links over time, and all agents can meet other agents and take link formation actions at every time step as long as their utility functions are not yet saturated. Based on such dynamic model, a distinguishing characteristic of a network is the time span over which agents keep forming links, i.e. how long would an agent keep searching for friends upon its arrival. Based on the definition of the utility function in (1) and (2), we know that there exists a time after which an agent stops forming links, which follows from our assumption of the concavity of the utility function. Moreover, the minimum number of friends that an agent makes in the network reflects the agent's gregariousness. The time horizon over which the agent forms this number of links is a function of all the exogenous parameters since it clearly depends on the meeting process, i.e. the time span over which an agent forms links is random as it depends on the types of agents that the agent meets over time and the history of the link formation decisions. For an agent i, the time span of link formation T_i is defined as

$$T_i \triangleq \inf\{t \in \mathbb{N} : a_i^\tau = 0, \forall \tau > t\} - i + 1. \qquad (5)$$

Since T_i is random, we characterize the time spent by an agent in the link formation process in terms of the expectation of

---

[2]Our definition for social capital follows that by Bourdieu in [35].

[3]While we assume concavity of the utility function, our analysis applies to any saturating function, e.g. the *sigmoid* function.

[4]For instance, in citation networks, the number of references cited in a paper is finite and corresponds to the number of papers the authors need to acquire knowledge, yet the number of citations on a specific paper can be arbitrary large.

[5]A detailed proof can be found in Appendix A.

$$a_i^t = \frac{1}{2}\,\mathrm{sgn}\left(v\left(\sum_{j\in\mathcal{N}_{i,t-1}^+}\alpha_{\theta_i\theta_j}+\alpha_{\theta_i\theta_{m_i(t)}}\right)-v\left(\sum_{j\in\mathcal{N}_{i,t-1}^+}\alpha_{\theta_i\theta_j}\right)-c\right)+\frac{1}{2}. \tag{2}$$

T_i. Note that T_i can be thought of as the *stopping time* of the meeting process $M_i(t)$. This can be easily proven by showing that the event $T_i = T$ only depends on the history of meetings and link formation decisions up to time T. We define the Expected Link Formation Time (ELFT) $\overline{T}_i$ as

$$\overline{T}_i = \mathbb{E}\left[T_i \,|\, i, \theta_i\right], \tag{6}$$

where the expectation is taken over the probability mass function (pmf) of T_i, which we denote by $f_{T_i}(T_i)$. In the following Theorem, we compute the ELFT for extreme cases of agents' homophily.

Theorem 1: (Homophily induces uncertainty) For an agent i born in an asymptotically large network ($i \to \infty$), if $h_k = 0, \forall k \in \Theta$, then the LFT for agent i and is equal to

$$T_i = L_{\theta_i}^*(\theta_i, 0)$$

almost surely, while if $h_k = 1, \forall k \in \Theta$, then the ELFT is given by

$$\overline{T}_i = \frac{1}{p_{\theta_i}} + \frac{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i\theta_i})}{\gamma p_{\theta_i} + (1-\gamma)}.$$

Proof: See Appendix B. ∎

Thus, when the agents are *not* homophilic, there is no uncertainty in the friendships acquisition process, and both the number of links and the link formation time are deterministic. This deterministic LFT is independent on the network, and only depends on the agent's gregariousness. That is, if $h_k = 0, \forall k \in \Theta$, then an agent's journey in the network is controlled by the agent itself and how it values friendships, and not by the network structure or the actions of others. If agents value friendships more, i.e. are more gregarious, then they will spend more time making friends, yet this time is deterministic and only depends on parameters that are determined by the agent and not the network. On the other hand, if agents are extremely homophilic, then the agent's journey in the network will entail more stochasticity, i.e. agents are less certain about the time needed to form links since they can meet different-type agents with which they do not form any links. It is clear from Theorem 1 that the ELFT of extremely homophilic agents depends on the type distribution and opportunism, in addition to gregariousness. We emphasize these dependencies in the following corollary.

Corollary 1: (Gregarious agents and minorities search for friends longer, opportunistic agents search shorter) If $h_k = 1, \forall k \in \Theta$, then for an agent i, the ELFT is:

- a monotonically increasing function agent i's gregariousness $L_{\theta_i}^*(\theta_i, 0)$.
- a monotonically decreasing function of p_{θ_i}, and
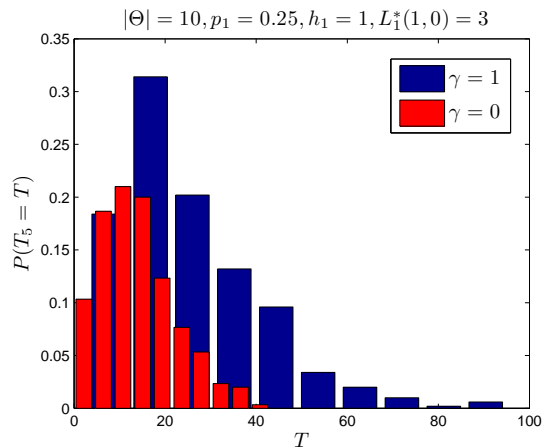- a monotonically increasing function of γ.

Fig. 2: Stochastic dominance of the LFT with decreasing structural opportunism.

Proof: See Appendix C. ∎

Corollary 1 says that the ELFT is monotonic in gregariousness, which is intuitive since the more friends an agent is willing to make, the longer it takes to find those friends. Moreover, agents belonging to minorities are expected to spend more time in the link formation process. This is again intuitive since when the fraction of similar-type agents in the population is small, each agent would need to meet a longer sequence of agents in order to find similar-type friends. Finally, the ELFT decreases as structural opportunism increases. This is because once the agent becomes attached to a network component of similar-type agents, it is then better to be opportunistic and keep meeting the friends of friends who are guaranteed to be similar-type agents, rather than meeting strangers with uncertain types.

Note that the results in Theorem 1 are concerned with extreme cases of homophily, i.e. $h_k = 0$ and $h_k = 1$. Computing the exact ELFT for an arbitrary exogenous homophily index is not mathematically tractable due to the combinatorial nature of the agents' interactions. However, in the following Theorem, we generalize Theorem 1 and Corollary 1 using *stochastic ordering*[6].

Theorem 2: (Stochastic ordering of the LFT statistics with respect to exogenous parameters) In an asymptotically large network, for any agent i with an exogenous parameters tuple $(p_{\theta_i}, h_{\theta_i}, \gamma, L_{\theta_i}^*(\theta_i, 0))$, and a corresponding pmf of the LFT $f_{T_i}(T_i)$, we have:

- If $\tilde{p}_{\theta_i} > p_{\theta_i}$ and all other exogenous parameters are the

[6]The definition for stochastic dominance can be found in section 4.5.5 in [8].

same, then $f_{T_i}(T_i)$ *first-order stochastically dominates* $\tilde{f}_{T_i}(T_i)$.

- If $\tilde{h}_{\theta_i} > h_{\theta_i}$ and all other parameters are the same, then $\tilde{f}_{T_i}(T_i)$ first-order stochastically dominates $f_{T_i}(T_i)$.
- If $\tilde{\gamma} > \gamma$ and all other parameters are the same, then $\tilde{f}_{T_i}(T_i)$ first-order stochastically dominates $f_{T_i}(T_i)$.

Proof: See Appendix D. ∎

Theorem 2 generalizes Theorem 1 and Corollary 1 in the sense of stochastic ordering; the monotonicity of the ELFT as a function of the exogenous parameters, as well as the expectation of any other increasing function over the LFT pmf, follows the same behavior in Corollary 1. Note that while the ELFT can be shown to be increasing with an agent's gregariousness, the social gregariousness is actually coupled with the homophily index, thus one cannot change the gregariousness while fixing a homophily index. Fig. 2 shows that the pmf of the LFT for $\gamma = 1$ stochastically dominates that for $\gamma = 0$.

IV. POPULARITY ACQUISITION: HOW POPULAR CAN AN AGENT BECOME?

Each agent forms a finite number of links that satisfies its social gregariousness, but the number of links that an agent receives (which quantifies the agent's popularity) can be arbitrarily large depending on the agent's type, age, and position in the network. In this subsection, we characterize the evolution of the agents' popularity over time. While the ELFT measures the time span over which the agent forms links, we introduce another measure for the time span over which an agent attains a certain level of popularity, which we term the Expected Popularity Acquisition Time (EPAT). Let $T_i^p(d)$ be the *popularity acquisition time*, i.e. the time period over which an agent's indegree becomes d, i.e.

$$T_i^p(d) \triangleq \inf \left\{ t \in \mathbb{N} : \deg_i^-(\tau) \geq d, \forall \tau > t \right\} - i + 1. \quad (7)$$

Note that $T_i^p(d)$ can be thought of as the difference between the *hitting time* of the process $\left\{ \deg_i^-(t) \right\}_{t=i}^{\infty}$ and the birth date of agent i. Since such period is random, we define the *expected popularity acquisition time* (EPAT) $\overline{T}_i^p(d)$ as

$$\overline{T}_i^p(d) = \mathbb{E}\left[T_i^p(d) \,|\, i, \theta_i \right]. \quad (8)$$

The popularity of agent i at time t is simply given by $\deg_i^-(t)$. We say that the *popularity growth rate* is $O(g(t))$ if $\lim_{t \to \infty} \left| \frac{\mathbb{E}\left\{ \deg_i^-(t) \right\}}{g(t)} \right| = b$, where b is a positive constant. Finally, we define the *popularity crossover time* $T_c^p(i, j, \theta_i, \theta_j, \gamma, \gamma')$, as the time at which the expected popularity of an agent j of type θ_j in a network realization with structural opportunism γ' exceeds the popularity of an agent i of type θ_i in a network realization with structural opportunism γ, i.e.

$$T_c^p(i, j, \theta_i, \theta_j, \gamma, \gamma') \triangleq$$

$$\inf \left\{ t \in \mathbb{N} : \mathbb{E}\left\{ \deg_i^-(\tau) \,|\, \gamma \right\} < \mathbb{E}\left\{ \deg_j^-(\tau) \,\Big|\, \gamma' \right\}, \forall \tau > t \right\}. \quad (9)$$

When $\gamma = \gamma'$, we simply denote the popularity crossover time by $T_c^p(i, j, \theta_i, \theta_j)$. Next, in the following, we define a new notion of *doubly preferential meeting processes*, which plays a central role in the popularity evolution process.

Definition 1: (*Doubly preferential meeting process*) We say that the meeting process $M_i(t)$ is *doubly preferential* if

$$\mathbb{P}\left(m_i(t) = j \,\Big|\, i \leq t \leq i + T_i - 1, \deg_j^-(i) \right) = y_{ij}\left(\deg_j^-(i) \right),$$

where $y_{ij}(x)$ is a linear function of x, and $y_{ij}(x) > y_{ik}(x), \forall x$ if and only if $p_{\theta_j} r_{\theta_i}(\theta_j, \infty) > p_{\theta_k} r_{\theta_i}(\theta_k, \infty)$. ∎

Thus, a doubly preferential meeting process is a process that leads agent i to meet an agent j with a probability proportional to both the current popularity level of agent j, and the maximum excess representation of type-θ_j agents in the friends of type-θ_i agents. The meeting process is *doubly preferential* because it gives an advantage to the more popular agents, and in addition gives an advantage to similar-type agents. For an extremely homophilic agent i ($h_{\theta_i} = 1$), we have $y_{ij}(x) = 0, \forall x, \theta_i \neq \theta_j$. It is worth mentioning that since doubly preferential meetings allows similar-type agents to meet with higher probability, it then captures what Mayhew calls "structuralist" homophily effects in [10], and what Kossinets and Watts refer to as "induced homophily" in [11], together with the linear preferential attachment growth model. In the following Lemma, we show that structural opportunism and homophily promotes doubly preferential meeting processes.

Lemma 1: (*Structural opportunism and homophily promote doubly preferential meetings*) For any network with $\gamma \in [0, 1)$ and $h_k \in (0, 1], \forall k \in \Theta$, the meeting process of any agent i is doubly preferential. ∎

In the following Theorem, we show that *preferential attachment* in the popularity evolution process emerges over time as a result of the doubly preferential meeting processes, which in turn results from structural opportunism.

Theorem 3: (*Emergence of preferential attachment due to structural opportunism*) For any network with $\gamma \in [0, 1)$, preferential attachment emerges over time and we have

$$\mathbb{P}\left(A^t(i, j) = 1 \,\big|\, \deg_j^-(i), i \leq t \leq i + T_i - 1 \right) \propto y_{ij}\left(\deg_j^-(i) \right).$$
∎

This Theorem says that if agents are opportunistic, then the popularity of each agent exhibits an *accumulated advantage* pattern where the *popular gets more popular*. Unlike [13] [14] [19] [36], in our model preferential attachment emerges endogenously over the link formation time as a result of the meeting process and the actions of the agents rather than being a probabilistic link formation rule that specifies a one-shot linking behavior. Moreover, the probability that agent i links to agent j depends not only on the popularity of agent j, but also on the gregariousness of agent i, and the types of agents i and j. Next, in the following Theorem we study the impact of the exogenous network parameters on the agents' popularity growth rates.

Theorem 4: (*Popularity growth in non-homophilic societies*) For a network with $h_k = 0, \forall k \in \Theta$, the popularity growth rates are given as follows:

- For $\gamma = 0$, the popularity of an agent i grows *sublinearly* with time, i.e. $\mathbb{E}\left\{\deg_i^-(t)\right\}$ is $O\left(t^{\frac{\bar{L}-1}{\bar{L}}}\right)$, where $\bar{L} = \sum_{k \in \Theta} p_k L_k^*(k, 0)$, and the EPAT grows super-linearly with the degree of popularity, i.e. $T_i^p(d)$ is $O\left(d^{\frac{\bar{L}}{\bar{L}-1}}\right)$.
- For $\gamma = 1$, the popularity of an agent i grows *logarithmically* with time, i.e. $\mathbb{E}\left\{\deg_i^-(t)\right\}$ is $O\left(\bar{L}\log(t)\right)$, and the EPAT grows exponentially with the degree of popularity, i.e. $T_i^p(d)$ is $O\left(e^{\frac{d}{\bar{L}}}\right)$. ∎

This Theorem demonstrates the impact of opportunism and gregariousness on popularity accumulation. The popularity of opportunistic agents grows sublinearly with time, and the sublinearity exponent depends on the "average" gregariousness of all types of agents in the society. On the other hand, if agents are not opportunistic, their popularity will grow only logarithmically with time. Thus, opportunism plays a fundamental role in determining the popularity growth rate. It is due to the emerging preferential attachment (which is promoted by agents' opportunism) that the popularity follows a sublinear growth over time. However, since the society is non-homophilic, the meeting process is not doubly preferential and the growth rates of all agents are the same. In fact, it is only the agent's age in the network that decide its expected popularity level as we show in the next Corollary.

Corollary 2: (Superiority of older agents in non-homophilic societies) For a network with $h_k = 0, \forall k \in \Theta$, $T_c^p(i, j, \theta_i, \theta_j) = \infty, \forall j > i$. That is, the expected popularity $\mathbb{E}\left\{\deg_i^-(t)\right\}$ of an agent i at any time step dominates the expected popularity of all younger agents. ∎

This Corollary says that popularity crossover does not occur in non-homophilic societies. This is a direct consequence of agents being indifferent to other agents' types, thus it is only the birth dates that distinguishes the agents. In the following Theorem, we compute the popularity crossover time for the same agent for $\gamma = 0$ and $\gamma = 1$, and show that opportunism leads to popularity gains in the long-run.

Theorem 5: (Opportunism is good in the long-run) For a network with $h_k = 0, \forall k \in \Theta$, the popularity crossover time $T_c^p(i, i, \theta_i, \theta_i, \gamma, \gamma')$ is always finite and can be approximated by

$$T_c^p(i, i, \theta_i, \theta_i, 1, 0) \approx i \times \left(-\bar{L}\, \mathcal{W}_{-1}\left(\frac{1}{\bar{L}}e^{\frac{-1}{\bar{L}}}\right)\right)^{\frac{\bar{L}}{\bar{L}-1}},$$

where $\mathcal{W}_{-1}(.)$ is the lower branch of the Lambert W function [37]. ∎

The popularity crossover time increases linearly with the agent's birth date, and grows as $O\left(\bar{L}\log(\bar{L})\right)$ with the agents' average gregariousness. Thus, younger and more gregarious agents need to wait longer to harvest the popularity gains attained by opportunism. Fig. 3 shows the expected popularity growth over time for agents born at $t = 10$, 20, and 30. Solid lines are the logarithmically growing popularity if agents are not opportunistic, while dashed lines correspond to the sublinearly growing expected popularity for the opportunistic agents, and it can be seen that a finite crossover time exists for all such agents.

In the results above, we focused on non-homophilic societies, where meetings are not doubly preferential and different types go through the same popularity evolution process. In the following Theorem, we evaluate the popularity growth rates in homophilic societies.

Theorem 6: (Popularity growth in homophilic societies) For a network with $h_k = 1, \forall k \in \Theta$, the popularity growth rates are given as follows:
- For $\gamma = 0$, the popularity of an agent i grows *sublinearly* with time, i.e. $\mathbb{E}\left\{\deg_i^-(t)\right\}$ is $O\left(t^b\right)$, and the EPAT grows super-linearly with the degree of popularity, i.e. $T_i^p(d)$ is $O\left(d^{\frac{1}{b}}\right)$, where $b = \frac{L_{\theta_i}^*(\theta_i, 0) - 1}{L_{\theta_i}^*(\theta_i, 0)}$.
- For $\gamma = 1$, the popularity of an agent i grows *logarithmically* with time, i.e. $\mathbb{E}\left\{\deg_i^-(t)\right\}$ is $O\left(L_{\theta_i}^*(\theta_i, 0)\log(t)\right)$, and the EPAT grows exponentially with the degree of popularity, i.e. $T_i^p(d)$ is $O\left(e^{bd}\right)$, where $b = \frac{1}{L_{\theta_i}^*(\theta_i, 0)}$. ∎

It is clear that for homophilic societies, the popularity growth rate are the same as those computed in Theorem 4, yet the sublinearity exponent for opportunistic agents' popularity growth, and the scaling factor for the logarithmic popularity growth of non-opportunistic agents depend on the gregariousness of each type of agents. Thus, agents that are more gregarious experience faster growth rates, which implies that doubly preferential meeting process promotes the popularity of *older* and more *gregarious* agents by the effects of preferential attachment and homophily respectively as we show in the following Corollary.

Corollary 3: (Younger and more gregarious agents are more popular than older and less gregarious agents) For a network with $h_k = 1, \forall k \in \Theta$, $T_c^p(i, j, \theta_i, \theta_j) < \infty, \forall j > i$ if and only if $L_{\theta_j}^*(\theta_j, 0) > L_{\theta_i}^*(\theta_i, 0)$. ∎

Thus, the fraction of type-k agents in the society does not affect its chances of getting popular. In fact, it is the gregariousness of type-k agents, in addition to the structural opportunism, that control the popularity evolution. Structural opportunism leads to preferential attachment, which gives a cumulative advantage to older agents to get more popular, while homophily creates a doubly preferential meeting process, which together with the heterogeneity in the agent's gregariousness can promote the popularity of certain types of agents. The net effect can be that, unlike the case of non-homophilic societies, a younger agent can on average get more popular than an older agent if the younger agent belongs to a more gregarious type. Fig. 4 shows that the popularity of a younger, but more gregarious agent born at $t = 30$ exceeds that of an older agent born at $t = 20$. The results of Theorem 6 can be used to interpret the interesting empirical results on citation networks in [38], where it has been shown that there is a correlation between the number of references per paper and the total number of citations in a certain scientific field. We quote the following conclusion from the report in [39], which is based on a statistical analysis of Thomson Reuters Essential Science Indicators database: "*One might think that the number of papers published or the population of researchers in a field are the predominant factors*
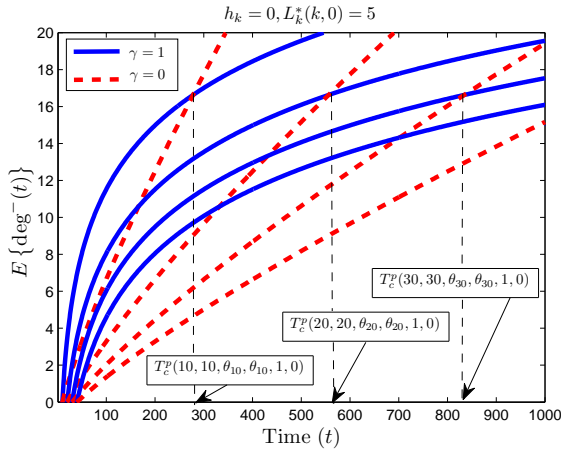
Fig. 3: Popularity crossover times for agents with and without structural opportunism. Agents are born at $t = 10$, 20, and 30.
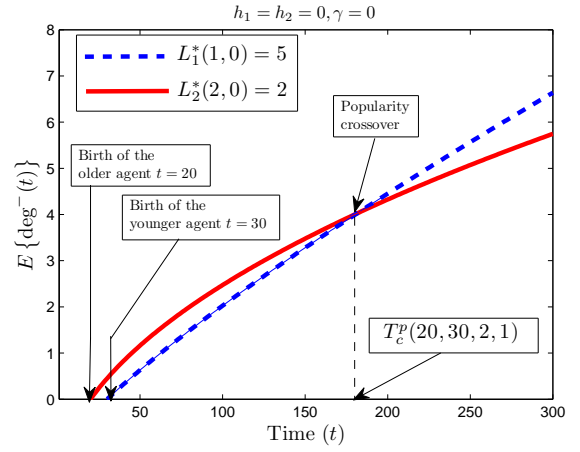

Fig. 4: Popularity crossover times in homophilic societies.

Fig. 5: Popularity crossover for molecular biology researchers with respect to mathematicians.

that influence the average rate of citation, but it is mostly the average number of references presented in papers of the field that determines the average citation rate." Such conclusion is in perfect agreement with Theorem 6 (and Corollary 3), where we predict that for the inherently homophilic citation networks, the popularity of researchers in different fields (total citation rate) is governed by their "gregariousness" (number of references per paper), and not by the type distribution (number of papers/researchers). We know from [39] that mathematics papers typically list few references, whereas those in molecular biology display extensive citations. Thus, molecular biologists are more "gregarious" than mathematicians, and one would expect that younger molecular biologists can, on average, get more popular than mathematicians. To illustrate this, we have collected the publicly available Google Scholar citation data for all molecular biologists who started publishing papers in the year 2000, and the citation data for mathematicians who started publishing in 1993. As predicted by Theorem 6 and Corollary 3, we show in Fig. 5[7] that a popularity crossover occurs for the "young" molecular biologists in the year 2006.

V. DISCUSSION AND FUTURE EXTENSIONS

A comprehensive theoretical model for the evolution of social networks was presented, and the associated analysis for agent-level link formation and acquisition was carried out. In this section, we discuss various uses and applications of our model. In particular, the model can be used, in addition to providing insights into the network evolution process, to carry out agent and network-level predictions, and designing policies that guide network formation.

A. Network Prediction

Our model can make useful predictions about the network structure and individuals popularity, time to acquire friendships etc. and how these change when the interacting agents

---

[7]Note that popularity growth in Fig. 5 is super-linear due to the non-stationary arrival process and gregariousness, which we will incorporate in our future extensions for the model.
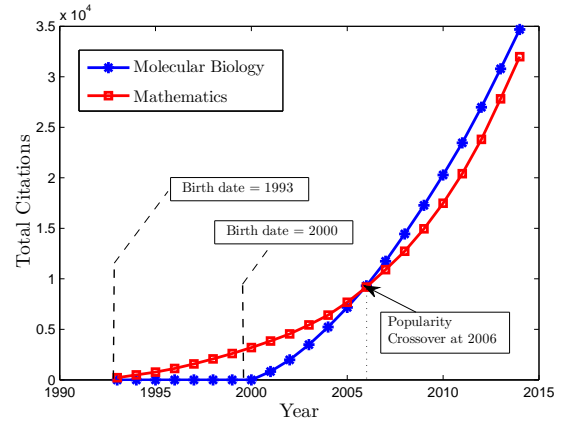
exhibit different characteristics. Moreover, our model predicts the emergence of preferential attachment, which is widely observed in many real-world networks [13], as a result of the agents' dynamic meeting process and linking actions. In contrast, data driven approaches [4]-[7] [29]-[30] cannot provide such network characterizations and predictions unless they have access to an abundance of relevant data. Moreover, the proposed microfoundational model enables us to determine what networks will form and how will they evolve when the agents characteristics are different and to understand what would be different if the agents would have different characteristics. Such detailed comparisons, analysis and counterfactuals are not possible based on data-driven approaches because this would require access to enormous amounts of data and, even more importantly, access to networks that cannot be monitored and may not even exist (yet). For instance, Leskovec *et. al* [6] characterized the friendship acquisition time, where a parameterized model is constructed with the aid of a large data set of temporal *node* (agent) arrivals and *edge* (link) creation times for Linkedin, Yahoo! Answers, and Flickr. The *time gap* between the creation of two links by the same node

is shown to fit an exponential distribution, and estimates for the exponential distribution parameter for different networks are provided. While [6] estimates the exponential distribution parameters for different networks, it implicitly assumes that all types of agents go through the same experience, and in addition cannot explain why different networks have different time gap statistics, and what will happen if different parameters are changed. In contrast, by calibrating the values of the exogenous parameters, our model can explain why the LFT of agents in different networks would differ, how is it affected by the exogenous parameters, and can in addition handle counterfactuals.

B. Policy design

Our model can be used to not only carry out link predictions, but it can also be used to analyze and design new policies that can modify and guide the formation and evolution of networks. An example for a policy is to influence the agents' meeting process at every time step given the observed step graph at the previous time step. This corresponds to the *link recommendation* problem [40], i.e. suggesting friends for the agent over time. Such process of guiding link creation are of interest to network planners and designers since in many OSN, such as Facebook, the friends recommendation system is responsible for a significant fraction of link creations. The policy design problem can have different objectives such as: controlling the level of popularity of different types of agents, minimizing the LFT, or controlling the community structure.

APPENDIX A
DERIVATION OF THE EXOGENOUS HOMOPHILY INDEX

The exogenous homophily index for type-k agents is given by

$$h_k = 1 - \frac{\sum_{m \neq k}^{|\Theta|} p_m \lim_{t \to \infty} r_{\theta_i}(m, t)}{1 - p_k}, \forall \theta_i = k,$$

thus, we need to evaluate the term $\lim_{t \to \infty} r_{\theta_i}(m, t)$, where

$$\lim_{t \to \infty} r_{\theta_i}(m, t) = \lim_{t \to \infty} \sup_{\mathcal{N}_{i,t}^+} \frac{\left| \mathcal{N}_{i,t}^{+,m} \right|}{\deg_i^+(t)}.$$

It is easy to see that the maximum excess representation of type-m agents can be evaluated by looking at all the possible set of friends that contains only type-m and type-k agents, i.e.

$$\lim_{t \to \infty} r_{\theta_i}(m, t) = \lim_{t \to \infty} \sup_{\mathcal{N}_{i,t}^+ = \mathcal{N}_{i,t}^{+,\theta_i} \cup \mathcal{N}_{i,t}^{+,m}} \frac{\left| \mathcal{N}_{i,t}^{+,m} \right|}{\deg_i^+(t)},$$

which is equivalent to

$$\lim_{t \to \infty} r_{\theta_i}(m, t) = 1 - \lim_{t \to \infty} \inf_{\mathcal{N}_{i,t}^+ = \mathcal{N}_{i,t}^{+,\theta_i} \cup \mathcal{N}_{i,t}^{+,m}} \frac{\left| \mathcal{N}_{i,t}^{+,\theta_i} \right|}{\deg_i^+(t)}.$$

The constraint $\mathcal{N}_{i,t}^+ = \mathcal{N}_{i,t}^{+,\theta_i} \cup \mathcal{N}_{i,t}^{+,m}$ implies that we restrict the links formed by agent i to agents of types θ_i and m only. Thus, h_{θ_i} can be computed by studying the sample paths for agent i where $\theta_{m_i^m(t)} \in \{\theta_i, m\}, \forall t \in \{i, i+1, \ldots, i+T_i-1\}$,

where $m_i^m(t)$ is a meeting process that restricts the types of agents met by agent i to types θ_i and m, and its corresponding sample space is $\mathcal{M}^m \subset \mathcal{M}$. Thus, we have that

$$\lim_{t \to \infty} \inf_{\mathcal{N}_{i,t}^+ = \mathcal{N}_{i,t}^{+,\theta_i} \cup \mathcal{N}_{i,t}^{+,m}} \frac{\left| \mathcal{N}_{i,t}^{+,\theta_i} \right|}{\deg_i^+(t)} =$$

$$\inf_{m_i^m(t) \in \mathcal{M}^m} \frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = \theta_i\}}}{\sum_{t=i}^{\infty} a_i^t},$$

which can be rewritten as

$$\inf_{m_i^m(t) \in \mathcal{M}^m} \frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = \theta_i\}}}{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = \theta_i\}} + \sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = m\}}},$$

which can be further reduced to

$$\inf_{m_i^m(t) \in \mathcal{M}^m} \frac{1}{1 + \frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = m\}}}{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = \theta_i\}}}}.$$

Due to the monotonicity and concavity of $v_i(x)$, it is clear that for two meeting processes $m_{i,1}^m(t)$ and $m_{i,2}^m(t)$, if $\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,1}^m(t) = m\}} > \sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,2}^m(t) = m\}}$, then $\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,1}^m(t) = \theta_i\}} < \sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,2}^m(t) = \theta_i\}}$, and $\frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,1}^m(t) = m\}}}{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,1}^m(t) = \theta_i\}}} > \frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,2}^m(t) = m\}}}{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_{i,2}^m(t) = \theta_i\}}}$.

Therefore, the term $\frac{1}{1 + \frac{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = m\}}}{\sum_{t=i}^{\infty} \mathbf{1}_{\{a_i^t = 1, m_i^m(t) = \theta_i\}}}}$ is minimized by maximizing the number of type-m agents in the friends of agent i. Now define $L_{\theta_i}^*(k, \alpha)$ as the maximum number of type-k links that can be established by agent i given that the agent i has initially as aggregate benefit of α. $L_{\theta_i}^*(k, \alpha)$ can be computed by taking the first derivative of the utility function with respect to type-k links, thus we have

$$L_{\theta_i}^*(k, \alpha) = \begin{cases} \lfloor L \rfloor & : v_i(L\alpha_{\theta_i k}) - v_i((L-1)\alpha_{\theta_i k}) < c \\ \lceil L \rceil & : v_i(L\alpha_{\theta_i k}) - v_i((L-1)\alpha_{\theta_i k}) > c \end{cases}$$

where

$$L = \frac{1}{\alpha_{\theta_i k}} \left(g_i^{-1} \left(\frac{c}{\alpha_{\theta_i k}} \right) - \alpha \right),$$

and $g_i(x) = \frac{\partial v_i(x)}{\partial x}$. Due to the concavity of $v_i(x)$, then $L_{\theta_i}^*(k, \alpha_1) > L_{\theta_i}^*(k, \alpha_2)$ if $\alpha_1 < \alpha_2$. Thus, the maximum number of type-m links in agent i's friends is $L_{\theta_i}^*(m, 0)$, and the corresponding number of type-θ_i links is $L_{\theta_i}^*(\theta_i, \alpha_{\theta_i m} L_{\theta_i}^*(m, 0))$. Consequently, the maximum excess representation of type-m links in agent i's friends is given by

$$\lim_{t \to \infty} r_{\theta_i}(m, t) = \frac{L_{\theta_i}^*(m, 0)}{L_{\theta_i}^*(m, 0) + L_{\theta_i}^*(\theta_i, \alpha_{\theta_i m} L_{\theta_i}^*(m, 0))}.$$

By substituting in (3), the exogenous homophily index for any agent i is then given by

$$h_{\theta_i} = \frac{1}{1 - p_{\theta_i}} \left(1 - \sum_{k=1}^{|\Theta|} p_k \frac{L_{\theta_i}^*(k, 0)}{L_{\theta_i}^*(k, 0) + L_{\theta_i}^*(\theta_i, \alpha_{\theta_i k} L_{\theta_i}^*(k, 0))} \right).$$

APPENDIX B
PROOF OF THEOREM 1

We start by studying the case when $h_{\theta_i} = 0$. When $h_{\theta_i} = 0$, then $\alpha_{\theta_i \theta_i} = \alpha_{\theta_i k} = \alpha, \forall k \in \Theta$. Therefore, at any date $t \geq i$, we have $\alpha_{\theta_i \theta_{m_i(t)}} = \alpha, \forall i \leq t$, and we have

$$\mathbb{P}\left(a_i^t = 1 \,\big|\, \theta_{m_i(t)}\right) = \begin{cases} 1 & : L_{\theta_i}^*(\theta_i, (t-i)\alpha) > 0 \\ 0 & : L_{\theta_i}^*(\theta_i, (t-i)\alpha) = 0 \end{cases}$$

Thus, agent i forms a link with all the agents it meets until the stopping time of $M_i(t)$, which in this case is equal to $L_{\theta_i}^*(\theta_i, 0)$, i.e.

$$\mathbb{P}\left(T_i = L_{\theta_i}^*(\theta_i, 0)\right) = 1, \forall i \in \mathbb{N}.$$

The LFT of agent i in the non-homophilic society is equal to $L_{\theta_i}^*(\theta_i, 0)$ almost surely since agent i forms a link with whatever type of agent it meets until its utility function is saturated. Now we focus on the case when $h_{\theta_i} = 1$. In this case, we have $L_{\theta_i}^*(k, 0) = L_{\theta_i}^*(\theta_i, 0)\,\delta(\theta_i, k), \theta_i, k \in \Theta$. Upon its arrival, agent i meets type-k agents picked uniformly at random from the network with probability p_k, and it forms its first link only if it meets a type-θ_i, i.e.

$$\lim_{i \to \infty} \mathbb{P}\left(a_i^t = 1 \,\big|\, \theta_{m_i(t)}, \deg_i^+(t) = 0\right) = \mathbf{1}_{\{\theta_{m_i(t)} = \theta_i\}},$$

therefore we have

$$\lim_{i \to \infty} \mathbb{P}\left(a_i^t = 1 \,\big|\, \deg_i^+(t) = 0\right) = p_{\theta_i}.$$

Let N_1 be the time elapsed until agent i forms its first link. It is clear that N_1 is a geometric random variable, and $\mathbb{E}\{N_1\} = \frac{1}{p_{\theta_i}}$. After agent i becomes attached to the network, the probability that it meets a type-θ_i agent is given by

$$\mathbb{P}\left(\theta_{m_i(t)} = \theta_i\right) = \gamma\,\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \bar{\mathcal{K}}_{i,t}\right) +$$
$$(1-\gamma)\,\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \mathcal{K}_{i,t}\right). \quad \text{(B.1)}$$

Since any type-k agent has $h_k = 1$, then all the friends of friends for an agent i has a type θ_i. Therefore $\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \mathcal{K}_{i,t}\right) = 1$. Moreover, we have

$$\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \bar{\mathcal{K}}_{i,t}\right) = \frac{|\mathcal{V}_{\theta_i}^t| - |\mathcal{K}_{i,t}| - \deg_i^+(t) - 1}{|\mathcal{V}^t| - |\mathcal{K}_{i,t}| - \deg_i^+(t) - 1}.$$

Note that the cardinality of the set $\mathcal{K}_{i,t}$ can be bounded as follows

$$\lim_{t \to \infty} |\mathcal{K}_{i,t}| \leq \left(L_{\theta_i}^*(\theta_i, 0)\right)^2.$$

Thus, for an asymptotically large network, we have

$$\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \bar{\mathcal{K}}_{i,t}\right) =$$

$$\lim_{t \to \infty} \frac{|\mathcal{V}_{\theta_i}^t| - |\mathcal{K}_{i,t}| - \deg_i^+(t) - 1}{|\mathcal{V}^t| - |\mathcal{K}_{i,t}| - \deg_i^+(t) - 1}, \quad \text{(B.2)}$$

which reduces to

$$\mathbb{P}\left(\theta_{m_i(t)} = \theta_i \,\big|\, m_i(t) \in \bar{\mathcal{K}}_{i,t}\right) =$$

$$\lim_{t \to \infty} \frac{p_{\theta_i} t - L_{\theta_i}^*(\theta_i, 0)\left(L_{\theta_i}^*(\theta_i, 0) + 1\right) - 1}{t - L_{\theta_i}^*(\theta_i, 0)\left(L_{\theta_i}^*(\theta_i, 0) + 1\right) - 1} = p_{\theta_i}. \quad \text{(B.3)}$$

Thus, the probability of forming a link at time step t is

$$\mathbb{P}\left(\theta_{m_i(t)} = \theta_i\right) = \gamma\,p_{\theta_i} + (1-\gamma). \quad \text{(B.4)}$$

After it forms its first link, agent i needs to form $L_{\theta_i}^*(\theta_i, 0) - 1 = L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})$ links. Let N_i be the number of time steps needed to form the ith link, where $\mathbb{E}[N_1] = \frac{1}{p_{\theta_i}}$. The time period needed to form the ith link N_i follows a geometric random variable, and the set of random variables $\{N_1, N_2, \ldots, N_{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})}\}$ are independent for a large network. Thus, we have

$$\mathbb{E}[N_j] = \frac{1}{\gamma\,p_{\theta_i} + (1-\gamma)}, \forall j \geq 2$$

and

$$\mathbb{E}\left[T_i = \sum_{j=1}^{L_{\theta_i}^*(\theta_i, 0)} N_j\right] = \mathbb{E}[N_1] + \mathbb{E}\left[\sum_{j=2}^{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})} N_j\right]$$
$$\text{(B.5)}$$
$$= \mathbb{E}[N_1] + \sum_{j=2}^{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})} \mathbb{E}[N_j]$$
$$= \frac{1}{p_{\theta_i}} + \frac{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})}{\gamma\,p_{\theta_i} + (1-\gamma)},$$

and the result of the Theorem follows.

APPENDIX C
PROOF OF COROLLARY 1

For $h_k = 1, \forall k \in \Theta$, we know that $\overline{T}_i = \frac{1}{p_{\theta_i}} + \frac{L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})}{\gamma p_{\theta_i} + (1-\gamma)}$. Since $L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i}) = L_{\theta_i}^*(\theta_i, 0) - 1$, then it follows that $\overline{T}_i$ is monotonically increasing in agent i's gregariousness, i.e. $\frac{\partial \overline{T}_i}{\partial L_{\theta_i}^*(\theta_i, 0)} = \frac{1}{\gamma p_{\theta_i} + (1-\gamma)} > 0$. Next, by taking the first derivative of the ELFT of agent i with respect to the fraction of type-θ_i agents, we have

$$\frac{\partial \overline{T}_i}{\partial p_{\theta_i}} = \frac{-1}{p_{\theta_i}^2} - \frac{\gamma\,L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})}{(\gamma p_{\theta_i} + (1-\gamma))^2} < 0.$$

Thus, agents belonging to minor types are expected to wait longer in the link formation process. Finally, the ELFT varries with the friendliness of the society in the following manner

$$\frac{\partial \overline{T}_i}{\partial \gamma} = \frac{-(p_{\theta_i} - 1)L_{\theta_i}^*(\theta_i, \alpha_{\theta_i \theta_i})}{(\gamma p_{\theta_i} + (1-\gamma))^2} > 0.$$

Thus, the ELFT of any agent is monotonically decreasing in the level of structural opportunism.

APPENDIX D
PROOF OF THEOREM 2

Before studying the comparative statics of the evolving network, we first start by showing that the link formation process of each agent can be modeled as an absorbing Markov chain. For an agent i, we define the vector-valued stochastic process $\{\mathbf{S}_i^t\}_{t=i}^{\infty}$, where $\mathbf{S}_i^t \in \mathbb{Z}^{|\Theta| \times 1}$, and $\mathbf{S}_i^t = \left(\mathcal{N}_{i,t}^{+,1}, \mathcal{N}_{i,t}^{+,2}, \ldots, \mathcal{N}_{i,t}^{+,|\Theta|} \right)^T$. Note that for any $t \geq i$, we have $\mathbf{1}^T \mathbf{S}_i^t = \deg_i^+(t)$. Since each agent starts with no friends, we have $\mathbf{S}_i^i = \mathbf{0}^{|\Theta| \times 1}$, and since agents do not break links, we have $\mathbf{S}_i^{t+1} \succeq \mathbf{S}_i^t, \forall t \geq i$. Since the meeting process $M_i(t)$ has a finite stopping time T_i, then we have $\mathbf{S}_i^{t+1} = \mathbf{S}_i^t, \forall t \geq T_i$. Now we show that $\{\mathbf{S}_i^t\}_{t=i}^{\infty}$ is an absorbing Markov chain. Note that since each agent acquires at most one friend at each time step, then the transition probabilities should satisfy the following condition

$$\mathbb{P}\left(\mathbf{S}_i^{t+1} = \mathbf{s}_i^{t+1} \middle| \mathbf{S}_i^t = \mathbf{s}_i^t \right) > 0,$$

if and only if $\mathbf{1}^T \left(\mathbf{s}_i^{t+1} - \mathbf{s}_i^t \right) \in \{0,1\}, \forall t \geq i$, and $L_{\theta_i}^* (\theta_{m_i(t)}, \sum_{j=1}^{|\Theta|} \mathbf{s}_i^t(j) \alpha_{\theta_i j}) > 0$, where $\mathbf{s}_i^t(j)$ is the jth element of $\mathbf{s}_i^t$.

A pmf $f(x)$ *first order stochastically dominates* a pmf $\tilde{f}(x)$ if the following conditions hold (See Section 4.5.5 in [22])

- $\sum q(x)f(x) \geq \sum q(x)\tilde{f}(x)$, for all non-decreasing functions $q(.)$,
- $\sum_0^y f(x) \leq \sum_0^y \tilde{f}(x)$ for all y,
- $\sum_y^{\infty} f(x) \geq \sum_y^{\infty} \tilde{f}(x)$ for all y.

Now we study the comparative statics of the evolving network when changing each of the exogenous parameters. We start

REFERENCES

[1] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," *in Proc. of WWW '10*, pp. 591-600, 2010.
[2] M. Thelwall and K. Kousha, "ResearchGate: Disseminating, Communicating, and Measuring Scholarship?," *Journal of the Association for Information Science and Technology*, vol. 66, no. 5, May 2014.
[3] M. Skeels and J. Grudin, "When social networks cross boundaries: a case study of workplace use of facebook and linkedin," *in Proc. of ACM GROUP '09*, pp. 95-104, 2009.
[4] W. Shrum, N. H. Cheek, and S. Hunter, "Friendship in School: Gender and Racial Homophily," *Sociology of Education*, vol. 61, no. 4, pp. 227-239, Oct. 1988.
[5] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 703-712, 2012.
[6] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic Evolution of Social Networks," *in Proc. of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 462-470, 2008.
[7] L. M. A. Bettencourt and D. I. Kaiser, "Formation of Scientific Fields as a Universal Topological Transition," *arXiv:1504.00319v1*, Apr. 2015.
[8] M. O. Jackson and B. W. Rogers, "Meeting Strangers and Friends of Friends: How Random are Social Networks?," *American Economic Review*, Oct. 2006.
[9] M. McPherson, L. S. Lovin, and J. M. Cook," Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, vol. 27, mo. 1, pp. 415-44, 2001.
[10] B. H. Mayhew, "Structuralism versus Individualism: Part 1, Shadowboxing in the Dark." *Social Forces*, vol. 59, no. 2, 1980.
[11] G. Kossinets, D. J. Watts, "Origins of homophily in an evolving social network," *American Journal of Sociology*, vol. 115, no. 2, 2006.

[12] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, Aug. 2001.
[13] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509-512, Oct. 1999.
[14] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, Aug. 2001.
[15] A.-L. Barabasi, R. Albert, H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1, pp. 173-187, Oct. 1999.
[16] M. J. Newman, "The Structure and Function of Complex Networks," *Society for Industrial and Applied Mathematics (SIAM) Rev.*, vol. 45, no. 2, 2003.
[17] A. Vazquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Physical Review E*, vol. 67, May 2003.
[18] Y. Bramoulle, and B. W. Rogers, "Diversity and popularity in social networks," *Cahier de recherche/Working Paper*, vol. 9, 2009.
[19] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna, and D. Krioukov, "Popularity versus similarity in growing networks," *Nature*, no. 489, pp. 537-40, Sept. 2012.
[20] M. L. de Almeida, G. A. Mendes, G. M. Viswanathan, and L. R. da Silva, "Scale-free homophilic network", *The European Physical Journal B*, vol. 86, no. 6, 2013.
[21] P. Pin and B. W. Rogers, "Stochastic network formation and homophily," *To appear in The Oxford Handbook on the Economics of Networks*, Oct. 2013.
[22] M. O Jackson and A. Wolinsky, "A strategic model of social and economic networks," *Journal of economic theory*, vol. 71, no. 1, 1996.
[23] V. Bala, S. Goyal, "A noncooperative model of network formation," *Econometrica*, vol. 68, no. 5, pp. 1181-1229, Sept. 2000.
[24] S. Currarini, M. O. Jackson, and P. Pin, "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, vol. 77, no. 4, pp. 1003-045, Jul. 2009.
[25] S. Currarini and F. V. Redondo, "A Simple Model of Homophily in Social Networks," *University Ca Foscari of Venice, Dept. of Economics Research Paper Series*, vol. 24, Apr. 2013.
[26] S. Currariniy, M. O. Jackson, and P. Pin, "Identifying the Roles of Choice and Chance in Network Formation: Racial Biases in High School Friendships", *Proceedings of the National Academy of Sciences*, vol. 107, pp. 4857-4861, 2010.
[27] Y. Song and M. van der Schaar, "Dynamic Network Formation with Incomplete Information," *Economic Theory*, vol. 59, no. 2, 2015.
[28] Y. Zhang and M. van der Schaar, "Strategic Networks: Information Dissemination and Link Formation Among Self-interested Agents," *IEEE J. Sel. Areas Commun. - Special issue on Network Science*, vol. 31, no. 6, pp. 1115-1123, June 2013.
[29] L. Backstrom and J. Kleinberg, "Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook," *Proceedings of the 17th ACM conference on Computer supported cooperative work and social computing*, pp. 831-841, 2014.
[30] G. Kossinets, and D. J. Watts, "Empirical Analysis of an Evolving Social Network," *Science*, vol. 311, no. 88, 2006.
[31] L. Tang, H. Liu, and J. Zhang, "Identifying Evolving Groups in Dynamic Multimode Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 72-85, Jan. 2012.
[32] Z. Lu, Y. Wen, and G. Cao, "Information Diffusion in Mobile Social Networks: The Speed Perspective," *in Proc. IEEE INFOCOM*, 2014.
[33] I. M. Kloumann and J. M. Kleinberg, "Community membership identification from small seed sets," *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1366-1375, 2014.
[34] Y. Wang, H. Zang, and M. Faloutsos, "Inferring Cellular User Demographic Information Using Homophily on Call Graphs," *in Proc. of INFOCOM*, 2013.
[35] P. Bourdieu, "The forms of capital," *Handbook of Theory and Research for the Sociology of Education* (New York Greenwood), 1986.
[36] M. O. Jackson, "Social and Economic Networks," *Princeton University Press*, 2008.
[37] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, "On the Lambert W Function," *Advances In Computational Mathematics*, 1996.
[38] M. H. Biglu, "The influence of references per paper in the SCI to Impact Factors and the Matthew Effect," *Scientometrics*, vol 74, no. 3, 2008.

[39] https://www.timeshighereducation.co.uk/news/average-citation-rates-by-field-1998-2008/405789.article?storyCode=405789

[40] L. Backstrom and J. Leskovec,"Supervised Random Walks: Predicting and Recommending Links in Social Networks," *in Proc. of WSDM '11*, 2011.