A THEORY OF TWITTER

ROGAYEH DASTRANJ* AND ANKE KESSLER*
WORKING PAPER
OCTOBER 2014

Abstract. We develop a theoretical network-based model of the social network Twitter, formulating individual interaction as a dynamic game in which heterogenous agents choose a 'niche' (a subset of the type space) to tweet in, and whom to follow. Agents consume tweets close to their own types, and seek to maximize the number of their followers. Starting from any initial niche with an arbitrary length, we show that the dynamic Markov process converges to a niche with a finite maximum length, and this niche contains agent's own type. We also show that information does not diffuse as widely as one might expect: although many agents are directly or indirectly connected to each other, the news does not travel too far since agents strategically choose what news to tweet or retweet in accordance with their niche, i.e., they strategically *filter* information. We also discuss the stable networks that the dynamic process converges to in equilibrium and show that the star network is never stable if agents are similar enough, and the only stable network is the bidirectional full network. In contrast, when agents are further away in the type space, the star network is the only stable network.

1. Introduction

In the last decade, social media has become a predominant part of our societies, reshaping the way individuals communicate, interact, share information, coordinate activities, etc. Both Facebook and Twitter have grown exponentially since they were founded in 2004 and 2006 respectively. With more than 140 million active users and over 340 million messages posted per day, Twitter has become one of the most influential media for spreading and sharing breaking news, personal updates and spontaneous ideas [Wang et al (2013)]. In Facebook and Twitter, individuals are not just consumers

*Department of Economics, Simon Fraser University.
We would like to thank Matthew Jackson, Alexander Karaivanov, Sanjeev Goyal, and participants of the 11th Econometric Society World Congress (2015), and Canadian Economics Association meetings (2014) for helpful conversations, comments and feedback.

of information - as in traditional media - but they actively produce news and content which could potentially reach a broad group of people very fast. On the other hand, if they so choose, individuals could be exposed to a wide variety of opinions without too much effort; in Twitter, all they need to do is follow people with different viewpoints. By choosing who to follow, or who to be friend with, they choose what to be exposed to. This creates large degrees of homophily as well as segregation in these networks[1].

This paper presents a first attempt to formally model, and shed light on, what kinds of network configurations arise and how far information diffuses in Twitter, and to study whether or not the resulting networks are efficient. We develop a theoretical network-based model of Twitter that is meant to capture the main elements of the interaction on the network in a stylized way. Heterogenous users choose a 'niche' (a subset of the type space) to tweet in as well as whom to follow in each period. Agents consume tweets close to their own type and see to maximize the number of their followers. There might be several reasons for why individuals care about the number of their followers. One intuitive reason could be related to "status" and "conspicuous consumption": agents gain higher status or prestige as more agents follow them. Additionally, in Twitter, there is no official verification mechanism, so having more followers may signal that their tweets are more truthful, accurate or interesting. Another plausible reason might be that Twitter, and networked media in general, is extending the ability of users to create and receive personalized news streams [Hermida et al (2012)], hence, increasing users' influence on a large number of people. Therefore, one might summarize the utility that users get from content production as an increasing function of the number of people who get their messages.

We begin our analysis by characterizing the equilibria in the one-shot game, where agents simultaneously choose their niche and whom to follow. There is no retweeting in the one-shot game. We show even when the cost of following others is zero, there is an upper bound on the niche size as well as the number of followers and followings of each user, and depending on where their niches are, we find a range of equilibrium network configurations, from a connected network with completely connected components that exhibits high degrees of homophily, to a collection of bipartite networks with very little to no homophily at all.

In contrast to the static case, in the dynamic game, retweeting plays an important role. Retweeting can be seen as rebroadcasting a tweet, that a user received from his

[1]See, e.g. see Conover (2011) and Halberstam and Knight (2013)

followings, to his followers. It brings new people into a particular thread, inviting them to engage without directly addressing them [Boyd et al (2010)]. From network perspective however, retweets create the shortest path between agents in the network and provides an avenue to receive indirect benefits from others without directly following them. As we show, this leads to *filtering of information* by individuals: users do not retweet every tweet that they receive, as it could result in losing followers. Therefore, there is always some friction in diffusion of information through retweets.

Starting from any initial niche with an arbitrary length, we show that the dynamic Markov process converges to a niche with a maximum length, and this niche contains agent's own type. This is specifically true because users find it optimal to retweet in equilibrium. Retweets are a subset of tweets that users receive from their followings in the previous periods, and in equilibrium, agents follow others who tweet close their own type. Therefore, choosing niche further away from own type is equivalent to choosing not to retweet.

Additionally, we show that information does not diffuse as widely as one might expect: although many agents are directly or indirectly connected to each other, news does not travel too far since agents strategically choose what news to tweet or retweet in accordance with their niche. We also discuss the stable networks that the dynamic process converges to in equilibrium and show that the star network is never stable if agents are similar enough, and the only stable network is the bidirectional connected network with fully connected components. In contrast, when agents are further away in the type space, the star-like network is the only stable network. We conclude the paper by characterizing the efficient networks and show that the optimal number of followers from planners perspective is always greater than individuals' optimal followers in the one- shot game. However, in the dynamic game, depending on values of parameters in the model, planner's optimal number of followers maybe larger, equal or smaller than individuals' optimal choice.

Literature Review. There exists by now a fairly large literature on social networks in economics, as well as sociology, computer science, physics, etc. Much of the economics literature is reviewed in Jackson (2010) and Goyal (2009).

The paper in the network formation literature closest to our work is Bala and Goyal (2000). They present a network formation model where agents choose who to link to based on the trade off between the cost and benefits. They assume that the cost of link formation is only incurred by the individual who initiates the link (hence a noncooperative game) and study one-way and two-way flow of benefits. These assumptions are similar to our model, since in Twitter users choose whom to follow without their consent and enjoy the benefits of reading their tweets. However, in Twitter, agents not only consume information, but they also actively produce content and news, and therefore care about how many people follow them. Therefore, the link formation process is not just based on the benefits that agents receive from following others and the cost of maintaining the links, but also who agents follow, impacts who follows them. The benchmark model in Bala and Goyal (2000) is frictionless, i.e. they assume that if two agents are connected (even through several agents), the benefit that they receive from each other does not depend on the number of intermediaries. They show that in one-way flow model the only strict Nash networks are the wheel and the empty set. Their result hold for the dynamic game where agents either show inertia or randomize across their optimal strategies. Also, the wheel structure is the efficient network. However, when they introduce friction, it becomes much more difficult to characterize Nash equilibria.

In our model, in contrast, agents receive indirect benefit through retweets, which depends (endogenously) on whom agents follow and what tweets they receive in the previous periods. Therefore, there is always a degree of friction in Twitter, since individuals do not retweet everything that they receive from their followings. Additionally, characterization of the equilibrium networks in the dynamic game with friction is possible because of the connection between an agent's followings and followers through retweets.

The other important papers who study endogenous network formation are Watts (2003) and Jackson and Wolinsky (1996). Jackson and Wolinsky, study the relationship between stability and efficiency of networks when agents strategically choose who to form links with and show that the set of efficient and stable networks do not always coincide. Watts expands Jackson and Wolinsky's framework to dynamic setting and shows that the links formation process is path dependent and often converges to inefficient networks. In both of these models, both individuals' consent is needed for link formation, but any one of them could severe the link. Also, as in Bala and Goyal's model, individuals benefit directly and indirectly from being linked to others. In their dynamic game, in each period, a link is randomly (with uniform probability) identified. Agents are myopic and choose to form or severe a link based on that period's benefits.

In contrast, in our model, the direct and indirect benefit from others depends on the endogenous choice of users' niche.

More recently, there are a few papers studying the impact of social networks on homophily and segregation. The most relevant paper to our study is an empirical paper on measuring ideological homophily and segregation on Twitter. Using data from Twitter, Halberstam and Knight (2013) measure the levels of homophily and segregation among Twitter users. They find that their constructed network exhibits significant degrees of homophily, with links more likely to develop between individuals with similar ideological preferences, resulting in a much more segregated network than traditional media. They argue that social media may be a force for increasing isolation and ideological segregation in society.

Another paper that uses Twitter data to study political polarization is the study done by Conover et al (2011). They investigate how social media impacts the communication between different political communities and find that the network of political retweets exhibits a highly segregated partisan structure, with extremely limited connectivity between left- and right-leaning users. Surprisingly, though, they do not find the same pattern when they study the network of 'user-to-user mention', they see that this network has a single politically heterogeneous cluster of users in which ideologically-opposed individuals interact at a much higher rate compared to the network of retweets. These empirical results are in line with the predictions of our model which states that users tweet and follow others close to their types. This is particularly the case when individuals choose to retweet. On one hand, retweeting is profitable, since it creates shortest path in an environment that either following others is costly or finding worthy news and information is not easy, but on the other hand, by retweeting, users reveal their types, as they follow others who tweet close to their type. Therefore, we observe high degrees of homophily in equilibrium. Additionally, although we do not study 'user-to-user mentions' in our model, note that users can mention another individuals without necessarily following each other. So it is not too surprising that the 'user-to-user mention' network structure is not segregated at all: by mentioning another individuals, users could actually broadcast their opinion to a very different politically oriented community, who would not hear their message otherwise.

Tarbush and Teytelboym (2013) also develop a dynamic model of friending in the online social networks in which agents interact in overlapping social groups. By using mean-field approximation, they drive closed form analytical expressions for the distribution degree and homophily indices. They test and calibrate their model using Facebook data from 2005 and find empirical support for their model. Golub and Jackson (2012) also study how homophily affects communication and speed of learning in networks. An important distinction between these models and ours is that we do not assume homophily, and even in the one-shot game, we have equilibria without homophily. However, as discussed earlier, in the dynamic game homophily appears naturally in all the equilibria.

Finally, there is an extensive literature on Twitter in marketing, computer science and physics. Much of this literature, however, is concerned with patterns of diffusion of information and opinion without taking individual incentives into account.[2] For example, Kawamoto (2013) introduces a stochastic model of information diffusion using a random multiplicative process and justifies the model by observing the statistics of the multiplicative factors in the Twitter data. Also, Ko et al. (2014) present a mathematical model to extract the information sharing tendencies on Twitter in the 2012 presidential election in Korea.

The remainder of the paper is organized as following: in Section 2 we present the model. Section 3 and 4 respectively, are devoted to characterizing the equilibria of the one-shot as well as the dynamic game. The efficient networks from planner's perspective are presented in Section 5, followed by some concluding remarks.

2. The Model

There are n agents in the society $N = \{1, ..., n\}$ and each has a type $\theta_i \in T^1$, where $T^1 \equiv R^1/Z^1$ is a circle of circumference 1, i.e. $\theta_i \in [0, 1]$ where 0 and 1 are the same points on the circle. An individual's type θ_i is a random draw from the uniform i.i.d distribution of mass 1. This distribution function is known to all the agents. The parameter θ will determine an individual's preference for which information to consume and has many possible interpretations. For example, θ could represent agents' beliefs about the state of the world, their political views, their cultural heritage, their physical location, or simply, their 'interests' .

Each individual engaging in Twitter has two basic strategic decisions to make: (1) whom to follow and (2) what to tweet. We assume agents choose a niche $A_i = [\underline{x}_i, \overline{x}_i] \subseteq$

[2]See, e.g.,Yang et al (2010), Goel et al (2012), Morales et al (2012), Situngkir (2011), Rui et al (2013), Hui et al (2010), Libel et al (2009), Bakshy et al (2012), Wu et al (2011).

$[0,1]$ in the type space and the news that agents tweet arrive at random in this range: $x_i \in A_i$. Intuitively, this is because users usually tweet about a few topics rather than only one particular issue. Also news and information may arrive at random times, for example they may read an interesting article in a magazine or hear about another related topic on TV. Users choose their niche optimally in equilibrium.

The decision of whom to follow would result in forming directional links, since agents usually do not need permission to follow others.[3] If two agents follow each other, the link between them would be bi-directional. Throughout the paper we use the following notation to specify the direction of links:

$$\begin{cases} g_{ij} = 1 & \text{if } i \text{ follows } j \text{ and zero otherwise, } i \neq j \\ g_{ij} = g_{ji} = 1 & \text{if only if } j \text{ follows } i \text{ and } i \text{ follows } j,\ i \neq j \\ g_{ii} = 0 & \text{no link between an agent and himself (no loops)} \end{cases} \tag{2.1}$$

We say there is path of length $k+1$ from agent j to i, if there are agents $j_1, j_2, ..., j_k$ such that $g_{ij_1} = g_{j_1 j_2} = ... = g_{j_k j} = 1$.

There are two main assumptions governing individual behaviour that are maintained throughout:

Assumption **1.** *Individual utility from consuming a tweet is higher, the closer the content of the tweet is to their own type.*

Assumption **2.** *Individual utility is increasing in the number of followers.*

The first assumption states that agents would like to read tweets to maximize some function of $(|\theta_i - x_j|)$ where x_j is the (content of the) tweet – or retweet – by agent j. In other words, it posits that agents' benefit from reading a tweet has a consumption element, rather than stemming from pure informational content. This is a very natural assumption in the context of Twitter. While people on Twitter often follow at least one news media account to obtain current, and largely unfiltered, information about world or local events, the vast majority of the followings are composed of individuals whose tweets are limited to certain subjects. Following those individuals restricts the content of 'news' to certain areas of interest, for which assuming a consumption component seems plausible. Moreover, even in the case of pure informational content, individuals frequently exhibit a tendency to search for evidence which could confirm

[3]As exception are locked Twitter accounts, but those are relatively rare and we therefore ignore this possibility.

their own beliefs, rather than for evidence which could disconfirm it. This so-called *positive confirmation bias* has been well-documented in the psychology and behavioral economics literature.[4]

Formally, the utility that i receives from receiving a tweet from j, x_j, can be written as

$$V_i(\theta_i, x_j) = \begin{cases} -\beta|\theta_i - x_j| + \alpha & \text{if } \ g_{ij} = 1, \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

where β is a "tolerance parameter" representing an individuals' tolerance level: as β increases, agents get more disutility from hearing tweets that are further away from their own type. The parameter α is the maximum benefit that an individual would receive from a tweet. Assumption 1 implies that individuals will choose to follow people who tweet close to their own types, ceteris paribus. To rule out arbitrarily large followings and capture the congestion effect of "too much" information in one's news feed, however, followings come at a cost. Specifically, each person that an individual choses to follow imposes a (possibly very small) additional cost on that individual, which we take to be constant and equal to $c > 0$ for simplicity. The cost of following others is thus equal to $C(\sum_{j=1}^{N} g_{ij}) = c\sum_{j=1}^{N} g_{ij}$.

The second main behavioural assumption we make is that agents obtain utility from having more followers. As discussed in the Introduction, there is a multitude of reasons for why individuals would care about the number of their followers. One obvious argument, for example, would be be that by having more followers, users' tweets reach more people, increasing the utility they get from producing content as well as status or prestige. To fix ideas, we posit that this benefit is proportional to the number of followers, implying that we can write the utility from followers as $V(\sum_{j=1}^{N} g_{ji}) = v\sum_{j=1}^{N} g_{ji},\ v > 0$. A direct consequence of Assumption 2 is that an individual chooses his niche A_i, that is, the range of which to tweet in, in order to maximize the number of followers. For simplicity, we ignore any potential cost of tweeting (or retweeting) and instead assume that, once information randomly has arrived in one's niche, the individual can costlessly send a tweet (or retweet) containing the content of this information

[4] See, e.g., Jones and Snudgen (2001) and the references therein. As will become clear below, we could easily allow for pure informational interests without affecting the qualitative flavour of our results. What is important is that at least some followings are chosen on the basis of consumption of information and the benefit of this information being close to one's own type.

In what follows, we will first restrict ourselves to a one-shot game where agents simultaneously choose whom to follow, and what to tweet, in a single period. This will allow us to understand the basic workings of the model and build intuition. We subsequently introduce multiple periods, formalizing a dynamic game. In each case, a pure strategy for an agent (in each period) is to choose a niche and whom to follow: $s_i = \{A_i, \vec{g}_i = \{g_{ij} | g_{ij} = 1(\text{if i follows j}), g_{ij} = 0(\text{otherwise})\}\}$ and $s_i \in S_i \subset S$ where $S = \mathscr{G} \times T^1$ and $\mathscr{G}$ is the space of all networks with n agents.

3. Equilibria of Stage game: What if Twitter was only one period?!

In the one-shot game, agents simultaneously choose their niche and whom to follow. Each agent sends one tweet, which arrives at random (i.i.d.) in his niche A_i. Notice that in the one period game, there are no retweets. We also assume perfect information. The utility function of the stage game is:

$$U_i(\theta_i, A_i, A_{-i}, \vec{g}_i) = \sum_{j=1}^{N} g_{ij} \Big(E[-\beta|\theta_i - x_j| + \alpha] \Big) + v \sum_{j=1}^{N} g_{ji} - c \sum_{j=1}^{N} g_{ij} \tag{3.1}$$

Where $\vec{g}_i = \{j | g_{ij} = 1\}$, $A_i = [\underline{x}_i, \overline{x}_i]$, $\omega_i = |A_i|$ and $x_i \sim U_{iid}\ [\underline{x}_i, \overline{x}_i]$.

The equilibrium concept in this simultaneous move game is Nash equilibrium: agents play best response to their opponent's strategies. A strategy s_i is said to be best response to s_{-i} if

$$u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \qquad \forall s_i' \in S_i, s_{-i} \in S_{-i} \tag{3.2}$$

Lemma **1.** *In any symmetric equilibrium, we have*

a) *the optimal width of individuals' niches is* $\omega_i = \frac{2(\alpha - c)}{\beta} \leq \frac{2\alpha}{\beta}$, *where* $\omega_i = |A_i|$,
b) *the maximum number of agents that any individual followings and followers equals* $\frac{2(\alpha - c)}{\beta}$.

Proof. First note that the expected benefit that i receives from following j, if $\theta_i \leq \underline{x}_j$ or $\theta_i \geq \overline{x}_j$, is:

$$\begin{aligned}
E_{x_j\in[\overline{x}_j,\underline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] &= -\frac{\beta}{\omega_j}\int_{\underline{x}_j}^{\overline{x}_j} |\theta_i - x_j| dx_j + \alpha \\
&= -\frac{\beta}{\omega_j}\Big|\theta_i(\overline{x}_j - \underline{x}_j) - \frac{1}{2}(\overline{x}_j^2 - \underline{x}_j^2)\Big| + \alpha \\
&= -\beta\Big|\theta_i - \frac{1}{2}(\overline{x}_j + \underline{x}_j)\Big| + \alpha \qquad (3.3)
\end{aligned}$$

Where $\omega_j = (\overline{x}_j - \underline{x}_j)$, also note that $x_{j,\text{center}} = \frac{1}{2}(\overline{x}_j + \underline{x}_j)$ is the point in the middle of A_j. Therefore, we have

$$\begin{aligned}
E_{x_j\in[\overline{x}_j,\underline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] &= -\beta\Big|\theta_i - x_{j,\text{center}}\Big| + \alpha \\
\Rightarrow \quad E[B_j] \geq 0 \quad &\Leftrightarrow \quad \Big|\theta_i - x_{j,\text{center}}\Big| \leq \frac{\alpha}{\beta}
\end{aligned}$$

In other words, the last individual that would give i nonnegative expected benefit is an agent whose niche's middle point is at most $\frac{\alpha}{\beta}$ away from θ_i.

On the other hand, when $\underline{x}_j \leq \theta_i \leq \overline{x}_j$, then the above expectation is

$$\begin{aligned}
E_{x_j\in[\overline{x}_j,\underline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] &= -\frac{\beta}{\omega_j}\Big\{\int_{\underline{x}_j}^{\theta_i}(\theta_i - x_j)dx_j + \int_{\theta_i}^{\overline{x}_j}(x_j - \theta_i)dx_j\Big\} + \alpha \\
&= -\frac{\beta}{\omega_j}\Big[(\theta_i - x_{j,\text{center}})^2 + \frac{1}{4}\omega_j^2\Big] + \alpha \qquad (3.4)
\end{aligned}$$

Note that when $x_{j,\text{center}} = \theta_i$, $E[B_j] \geq 0 \;\Leftrightarrow\; \omega_j \leq \frac{4\alpha}{\beta}$. In other words, the maximum niche size for any agent is $\frac{4\alpha}{\beta}$ and if their niche is greater than this threshold, the expected benefit to any agent from following j is always negative.

Now take agent i, if he follows all the agents j whose niches are such that their middle points are at most $\frac{\bar{x}}{2}$ away from θ_i, his expected utility would be (when $c \neq 0$):

$$E[U_i] = \sum_j g_{ij} E_{x_j\in[\underline{x}_j,\overline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] + V(x_i(\omega_i)) - c\bar{x}$$

where $x_i(\omega_i)$ is the number of agents who follow i. Let $|\theta_i - x_{j,\text{center}}| = x_{ij}$, be the distance between agent i and the middle point of niche A_j. Note that in order to get non-negative expected benefit, $x_{ij} \leq \frac{\alpha}{\beta}$ and $\omega_j \leq \frac{4\alpha}{\beta}$, therefore

$$
\begin{aligned}
E[U_i] &= 2\int_0^{\frac{\bar{x}}{2}} \Big\{ -\frac{\beta}{\omega_j}\Big[(\theta_i - x_{j,\text{center}})^2 + \frac{1}{4}\omega_j^2\Big] + \alpha \Big\} dx_{ij} + V(x_i(\omega_i)) - c\bar{x} \\
&= 2\int_0^{\frac{\bar{x}}{2}} \Big\{ -\frac{\beta}{\omega_j}\Big[x_{ij}^2 + \frac{1}{4}\omega_j^2\Big] + \alpha \Big\} dx_{ij} + V(x_i(\omega_i)) - c\bar{x} \\
&= -\frac{\beta\bar{x}^3}{12\omega_j} - \frac{\beta\omega_j\bar{x}}{4} + \alpha\bar{x} + V(x_i(\omega_i)) - c\bar{x} \qquad (3.5)
\end{aligned}
$$

Therefore, the first order condition with respect to $\bar{x}$ is:

$$
\frac{\partial EU_i}{\partial \bar{x}} = -\frac{\beta\bar{x}^2}{4\omega_j} - \frac{\beta\omega_j}{4} + \alpha - c
$$

$$
\Rightarrow \quad \frac{\partial EU_i}{\partial \bar{x}} = 0 \quad \Leftrightarrow \quad \bar{x} = \sqrt{\frac{4(\alpha - c)\omega_j}{\beta} - \omega_j^2}
$$

Note that $\bar{x}$ is the total number of agents that i would follow to maximize his expected utility. These agents are such that the centre of their niches from θ_i is less than or equal to $\frac{\bar{x}}{2}$. Since the problem is symmetric, agent i knows that given $\omega_i \leq \frac{4\alpha}{\beta}$, only agents whose types are $\pm\frac{\bar{x}}{2}$ away from the centre of his niche ($x_{i,\text{center}}$) would follow him, i.e.

$$
x_i(\omega_i) = \sqrt{\frac{4(\alpha - c)\omega_i}{\beta} - \omega_i^2} \qquad (3.6)
$$

After taking the first order derivative of $x_i(\omega_i)$ with respect to ω_i, we see that the value of ω_i would maximize the number of i's followers is:

$$
\omega_i = \frac{2(\alpha - c)}{\beta} \qquad \Rightarrow \qquad x_i = \frac{2(\alpha - c)}{\beta} \qquad (3.7)
$$

which completes the proof. ■

The results in Lemma 1 states that both the niche size and the number of followings for each agent is a decreasing function of cost. When $c = 0$, niche size is $\frac{2\alpha}{\beta}$. Intuitively this makes sense, since when $|A_j| \leq \frac{2\alpha}{\beta}$, all the agents who are $\frac{\alpha}{\beta}$ away from the centre of the niche will get nonnegative expected benefit from j, i.e. the set of j's potential followers with this niche size, if $c = 0$, is $F_j = [x_{j,\text{center}} - \frac{\alpha}{\beta}, x_{j,\text{center}} + \frac{\alpha}{\beta}] = [\underline{x}_j, \overline{x}_j] = A_j$. However, when $|A_j| = \frac{2\alpha}{\beta} + \epsilon$, agent i who is $\frac{\alpha}{\beta}$ away from the centre of A_j, will get strictly negative expected utility from following j. Therefore, the set of followers decreases as $\omega_j > \frac{2\alpha}{\beta}$, in particular, when $|A_j| = \frac{4\alpha}{\beta}$, the only individual that would follow A_j is $\theta_i = x_{j,\text{center}}$. And if $|A_j| > \frac{4\alpha}{\beta}$, no one would follow j.

In summary, the set of followings, F_i, and followers, $H_{(i)i}$, for each agent i is:

$$F_i = [\theta_i - \frac{\alpha - c}{\beta}, \theta_i + \frac{\alpha - c}{\beta}]$$
$$H_i = [x_{i,\text{center}} - \frac{\alpha - c}{\beta}, x_{i,\text{center}} + \frac{\alpha - c}{\beta}] = A_i$$

If $c = 0$, then the maximum number of followings and followers is $\frac{2\alpha}{\beta}$, i.e. $F_i = [\theta_i - \frac{\alpha}{\beta}, \theta_i + \frac{\alpha}{\beta}]$. We will refer to this range, $[\theta_i - \frac{\alpha}{\beta}, \theta_i + \frac{\alpha}{\beta}]$ as the **tolerance range**, T_i, of agent i.

Proposition **1.** *The symmetric static equilibrium is characterized by:*

(1) *Individuals optimally choose niches of width $\omega_i = \frac{2(\alpha-c)}{\beta} \leq \frac{2\alpha}{\beta}$,*
(2) *Individuals follow others, whose centre of the niche falls within $\pm\frac{\alpha-c}{\beta}$ of their own type.*
(3) *Own niches do not necessarily coincide with the niches of their followings, i.e. θ_i may not be in i's own niche, A_i.*
(4) *The set of any agents' followers is the same as agents' own niche A_i, i.e. individuals are followed by $\frac{2(\alpha-c)}{\beta}$ other agents,*
(5) *The range of information that agent i receives from his followings, I_i, is greater or equal to his tolerance range, T_i, of agents. If $c = 0$, the range of information received is twice the tolerance range.*

Proof. (1), (2) and (4) directly follow from the previous lemma. (3) is an immediate consequence of the fact that consuming other tweets and choosing one's own niche are two independent actions, that are neither related in the individual utility function, nor strategically related through equilibrium interactions. To see (5), note that the last two agents that i follows are those whose niche centre points are $\theta_i + \frac{\alpha-c}{\beta}$ and $\theta_i - \frac{\alpha-c}{\beta}$. Therefore, the range of tweets that i would hear is $I_i = [\theta_i - \frac{2(\alpha-c)}{\beta}, \theta_i + \frac{2(\alpha-c)}{\beta}]$ which is twice the optimal niche size. When $c = 0$, $I_i = [\theta_i - \frac{2\alpha}{\beta}, \theta_i + \frac{2\alpha}{\beta}]$ which is twice his tolerance range. If $c = \alpha$, then the only person that i would follow is an agent with centre point of his niche exactly equal to θ_i, which means that he would only hear tweets in $[\theta_i - \frac{\alpha}{\beta}, \theta_i + \frac{\alpha}{\beta}]$, this is equal to i's tolerance range. ■

Note that (1) characterizes how wide an optimal niche should be, however, it does not specify where this niche should be: A_i could be centred around θ_i or could be somewhere else on the circle; in particular, (3) implies that multiple (symmetric) equilibria can emerge, i.e., there are many possibilities for 'where' agents could tweet that are consistent with equilibrium. As we will see shortly, this arbitrariness of niche locations and multiplicity of equilibria is an artefact of the one-shot game, and disappears once the dynamic nature of interactions, in particularly the fact that individuals can retweet the tweets of others, is taken into account. We conclude this section with some straightforward comparative static results:

Corollary **1.** *The equilibrium niche size and the number of users' followings and followers are decreasing in the tolerance parameter β and the cost of following c, and is increasing in the maximal utility from a tweet, α.*

From the discussion above, it is clear that depending on the position of the niche, different network configurations such as connected network with fully connected components as well as a collection of bi-partite networks, etc could be observed in equilibrium. For instance, assume there are only two types of users: 1, 2 (such that $|\theta_i - \theta_j| > \frac{\alpha}{\beta}\ \forall i \in 1,\ j \in 2$). Assuming symmetric equilibria, all $i \in 1$ could choose their niche to be centred around their own type and all $j \in 2$ also tweet around their own type. Therefore, agents of type 1 would follow each other and agents of type 2 would do the same and the network structure would be comprised of *two highly connected and isolated islands*. However, it is also possible that all the agents of type 1 tweet centred around type 2 and all the type 2 agents tweet around type 1. Then the resulting network is a *directed bi-partite* network that type 1's are connected to type 2 agents and vice versa, but there is no link from type 1's (or 2's) to themselves. The first network exhibits high degrees of homophily and segregation whereas there is no segregation in the second network.

4. Dynamic Game

In the dynamic game, there are infinite number of possibly very short periods, $t = 1, 2, ...$, and agents have discount factor δ. In each period, the one-shot simultaneous move game is repeated, where individuals can tweet from their niche (one randomly drawn tweet, as before) or retweet what they heard from their followings in the previous period. Given that the strategies in every period depends on the previous period, we define a Markov chain and study the Markov perfect equilibria (MPE).

In MPE, the history of the game is summarized in a state variable that is known to all the agents at the beginning of every period before they make their decisions. In this game:

- the state variable is the set of all tweets from the last period and the identity of the tweeter, $X_t = \{x_{i,t-1}\ \forall i : \text{set of all tweets in period } t-1\}$,
- the choice (control) variables are (1) what to tweet/retweet, (2) whom to follow: $s_{it} = \{A_{it}, \vec{g}_{it}\}$, with $s_t = s_t(X_t)$
- the state variable evolves according to $X_{t+1} = X_{t+1}(s_t)$.

Note that Markov strategies defined by the above state variable rule out grim trigger and other punishment strategies. Since the decisions in each period depends only on the state variable (all the tweets only from the last period) there are no punishment strategies such as "i would only follow j in period T, if j follows i for all $t < T$, otherwise, i would unfollow j", etc.

The Value function for the dynamic problem is:

$$V_i(X) = \max_{\{A_i, \vec{g}_i\}} \Big\{ EU_i(A_i, \vec{g}_i, X_i) + \delta V(X') \Big\} \tag{4.1}$$

$$s.t. \quad X' = X(A_i, \vec{g}_i)$$

$$EU_i = \sum_{j=1}^{N} g_{ij} E[-\beta|\theta_i - x_j| + \alpha](1 + \sum_{k=1}^{N} p_{jk} + \sum_{k=1}^{N}\sum_{l=1}^{N} p_{jk}p_{kl} + ...)$$

$$+ V(\sum_{j=1}^{N} g_{ji}) - C(\sum_{j=1}^{N} g_{ij}) \tag{4.2}$$

$$p_{ij} = \begin{cases} g_{ij}\frac{|A_i \bigcap A_j|}{|A_i|} & \text{if } i \text{ RT } j \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

The stage utility function is the same as in the static case with the difference that we have to account for the benefit from retweets. First, note that, in Twitter, users could retweet others whom they follow and it is not too difficult to see that they would do so as long as these tweets are in agent's own niche. This is because retweeting news outside of one's niche is equivalent to having a wider niche, and as argued in the on-shot game, when the niche width is greater than $\frac{2\alpha}{\beta}$, the set of user's potential followers decreases. This is also consistent with the assumptions that agents choose their niche and the tweets arrive in their niche. In other words, if i decides to retweet j, the probability that he receives a retweet-able tweet from j is increasing in the intersection between his own niche and j's niche. More explicitly, let $A_j = [\underline{x}_j, \overline{x}_j], A_k = [\underline{x}_k, \overline{x}_k]$:

(1) If $\underline{x}_j \leq \underline{x}_k \leq \overline{x}_j \leq \overline{x}_k$, then

$$
\begin{aligned}
p_{jk} &= g_{jk}\frac{|A_j \bigcap A_k|}{|A_j|} = \overline{x}_j - \underline{x}_k \\
&= g_{jk}\Big(\frac{1}{\omega_j}[\frac{(\omega_j+\omega_k)}{2} - x_{jk}]\Big) \qquad (4.4)
\end{aligned}
$$

where $\omega_{j,k} = |A_{j,k}|$ and $x_{jk} = |x_{j,\text{center}} - x_{k,\text{center}}|$,

(2) If $A_j \subseteq A_k$, then $p_{jk} = 1$,

(3) If $\underline{x}_k > \overline{x}_j$ or $\underline{x}_j > \overline{x}_k$ (i.e. $A_j \bigcap A_k = \emptyset$), then $p_{jk} = 0$.

Therefore, after imposing the symmetric equilibrium condition that $\omega_i = \omega \ \forall i$, we have:

$$
\begin{aligned}
p_{jk} &= g_{jk}(1 - \frac{x}{\omega}) \\
a_i &= 1 + \sum_{k=1}^{N} p_{jk} + \sum_{k=1}^{N}\sum_{l=1}^{N} p_{jk}p_{kl} + ...
\end{aligned}
$$

Assuming that in equilibrium j follows $\overline{x}_j$ other agents, the above sum in the continuous space is:

$$
\sum_{j=1}^{N} p_{ij} \quad \rightarrow \quad 2\int_0^{\frac{\overline{x}_i}{2}}(1 - \frac{x}{\omega})dx = \overline{x}_i - \frac{\overline{x}_i^2}{4\omega} \qquad (4.5)
$$

Similarly

$$
\begin{aligned}
\sum_{j=1}^{N}\sum_{l=1}^{N} p_{ij}p_{jl} \quad &\rightarrow \quad 2\int_0^{\frac{\overline{x}_i}{2}}(1 - \frac{x}{\omega})dx \times 2\int_0^{\frac{\overline{x}_j}{2}}(1 - \frac{x}{\omega})dx \\
&= (\overline{x}_i - \frac{\overline{x}_i^2}{4\omega})(\overline{x}_j - \frac{\overline{x}_j^2}{4\omega}) \qquad (4.6)
\end{aligned}
$$

In the symmetric equilibria $\overline{x}_i = \overline{x}_j = \overline{x}$ (i.e. each agent would follow $\overline{N}$ other agents), the infinite sum becomes:

$$
\begin{aligned}
a &= 1 + (\overline{x} - \frac{\overline{x}^2}{4\omega}) + (\overline{x} - \frac{\overline{x}^2}{4\omega})^2 + ... \\
&= \frac{4\omega}{\overline{x}^2 + 4\omega(1-\overline{x})} > 1 \qquad (4.7)
\end{aligned}
$$

This is true when $\overline{x} < 4\omega$ (note that both $\overline{x}$ and ω are less than one by assumption). Note that conditional on ω,

$$
\frac{da}{d\overline{x}} \geq 0 \quad \Leftrightarrow \quad \overline{x} \leq 2\omega \qquad (4.8)
$$

Using this intuition, the following lemma argues where in an environment that following others is costly, retweeting is always beneficial for all agents:

Lemma **2.** *When $c > 0$, retweeting is optimal for all agents.*

Proof. In an environment where following individuals is costly, retweets creates shortest path between agents, i.e. they could potentially hear tweets from agents who they don't follow directly. In other words, agents who retweet provide extra information and benefit to their followers. This increases their chance of being followed. More specifically, given that we have assume each agent only tweet once every period (but retweets as many as he can), the benefit from following i who retweets for any agent k is:

$$B_{ki}^{\text{rt}} = E[-\beta|\theta_k - x_i| + \alpha](1 + \sum_{j=1}^{N} p_{ij} + \sum_{j=1}^{N}\sum_{l=1}^{N} p_{ij}p_{jl} + ...) \tag{4.9}$$

The first term in the parenthesis is the direct benefit from i's own tweet and the other terms are the indirect benefit from retweet received from i's followings through i. It is easy to see that B_{ki}^{rt} is greater or equal to the benefit received from i when he does not retweet:

$$B_{ki}^{\text{rt}} \geq B_{ki}^{\text{no rt}} = E[-\beta|\theta_k - x_i| + \alpha] \tag{4.10}$$

More formally, assume that we are in an equilibrium such that each agent is following $m = \frac{2(\alpha-c)}{\beta}$ agents (who tweet close to their own types). Starting from an equilibrium that no one retweets, assume that with some positive probability i deviates and retweets tweets that he received in the previous period from his following in period t (as long as these retweets are in his niche). Therefore, the benefit that he provides for his followers is greater than other agents. Now, if k was following j such that $\theta_j = \theta_i \pm \epsilon$, then k would unfollow j and follow i in this period, since k would hear both i's tweet as well as j's tweet from i as retweets with a very high probability (in fact, i retweets all of his followings with a probability proportional to the intersection between their niches), while he is only incurring the cost of following one agent. In other words, all agents who are close to i would lose all their followers to i in the next period. Therefore, by this deviation, i increases his probability of being followed. Hence, this is a profitable deviation for i and by symmetry all agents finds it beneficial to deviate and retweet as well. ■

Intuitively, in a costly environment, retweeting provides a costless way of transmitting information through creating a shortest path. Retweeting provides several benefits to agents: by retweeting one of his followings, i not only spreads that particular tweet to his followers, but he also introduces his followings to his followers, increasing the probability that the agent who created the original tweet is followed by i's followers (who may not know him before). However, the important point is that, by retweeting, each agent creates a shortest path from and to other agents, this increases his own probability of being followed in the first place.

LEMMA **3.** *The optimal niche width and the number of agents' followings in the dynamic game is greater than the one-shot game.*

Proof. Recall that the optimal niche width and number of followings in the one-shot game were $\frac{2(\alpha-c)}{\beta}$. Now starting from this equilibrium, assume that j is the last agent that i follows to his left. If i deviates and follows the marginal agent $\bar{j}$ such that $\theta_{\bar{j}} = \theta_j - \epsilon$, then he has more tweets to retweet in the next period than other users and therefore, as before, all individuals who were following k such that $\theta_k = \theta_i - \epsilon$, would unfollow k and follow i. This change in i's utility is strictly positive due to envelope theorem. Since as $\epsilon \to 0$, the marginal cost of following the marginal agent is equal to the marginal benefit at $x_i = \frac{2(\alpha-c)}{\beta}$, but there is positive increase in the number of i's followers.

Formally, starting from a MPE, where agents niche is such that $A_i \subset I_i$, and they follow $\theta_j \in F_i = [\underline{\theta}_i, \overline{\theta}_i] \subseteq [\theta_i - \frac{\alpha}{\beta}, \theta_i + \frac{\alpha}{\beta}]\ \forall i$. Now assume i deviates and follows j such that $\theta_j = \underline{\theta}_i - \epsilon$. Note that by following j, i does not loose any followers, since he always filters which tweets to retweet in the following periods. However now, since he is following a marginal agent (even at a marginal cost), he could retweet j's tweets with probability:

$$\begin{aligned} p_{ij} &= \frac{|A_i \bigcap A_j|}{|A_i|} \\ &= 1 - \frac{x_{ij}}{\omega} \geq 0 \quad \Leftrightarrow \quad x_{ij} \leq \omega \end{aligned} \tag{4.11}$$

Put differently, as long as the distance between i and j is less than the niche width, i could retweet j with positive probability. Therefore, the expected utility of i from following all the agents in $[\underline{\theta}_i, \overline{\theta}_i]$ (before deviation) is:

$$U_i = \frac{1}{1-\delta}\Big\{\int_{\underline{\theta}_i}^{\overline{\theta}_i}(aE[-\beta|\theta_i - x_j| + \alpha])d\theta_j + V(\sum_j g_{ji}) - C(\bar{N})\Big\} \tag{4.12}$$

where $\bar{N} = \sum_j g_{ij} = (\overline{\theta}_i - \underline{\theta}_i)$ is number of agents that i follows, and a is the additional benefit from retweets (equation 4.7):

$$a = \frac{4\omega}{\overline{x}^2 + 4\omega(1-\overline{x})} > 1$$

where conditional on ω,

$$\frac{da}{d\overline{x}} \geq 0 \quad \Leftrightarrow \quad \overline{x} \leq 2\omega$$

This means that starting from a symmetric equilibrium where agents follow $\overline{x}$ agents with niche size ω, where $\overline{x} \leq 2\omega$, if i deviates and increase the number of his followings from $\overline{x} \to \overline{x} + \epsilon$, then there is strictly positive increase in the number of his followers because of the extra benefit (more retweets) he is providing that no one else is (all agents close to i (i.e. $\theta_i + \epsilon$) are going to lose their followers to agent i):

$$(1-\delta)dU_i|_{\bar{x}} = E[-\beta|\theta_i - x_j| + \alpha]d\theta_j + v\epsilon - cd\theta_j \tag{4.13}$$

Using the envelope theorem, $dU_i|_{\bar{x}} = v\epsilon > 0$. Hence, it is a profitable deviation to follow more agents than the one-shot game, regardless of the cost of following others.

The same argument is true when i deviates from the one-shot equilibrium niche and increases it by ϵ. This is because by doing so, he increases the probability of retweets (since he increases the intersection of his niche with his followings' niche) and hence, in expectation, increases his number of followers. ■

Proposition **2.** *Assuming uniform iid distribution of types, if $v > c$, then in every symmetric MPE:*

a) *The optimal width of agent's niche is $\omega_i \leq \frac{2\alpha}{\beta}$,*
b) *Individuals niche includes their own type: $\theta_i \in A_i$,*
c) *The optimal (maximum) number of followers and followings of any individual is equal to 2ω (which exceeds the static case) and these two sets coincide.*

Proof. Let us start from a MPE where agents do not tweet around their own type. However, all agents always follow others who tweet close to their type: $F_i \subseteq [\theta_i - \frac{\alpha}{\beta}, \theta_i + \frac{\alpha}{\beta}]$ $\forall i$, therefore, all the tweets that i receives from his followings are maximum (when $c = 0$) $\frac{2\alpha}{\beta}$ away from θ_i:

$$x_j \in [\theta_i - \frac{2\alpha}{\beta}, \theta_i + \frac{2\alpha}{\beta}] = I_i$$

Where I_i is the information set of agent i. If, in this MPE, A_i is such that $x_j \notin A_i \; \forall x_j$, then no one retweets (since by assumption agents only retweets tweets that are in A_i). However, if i deviates and chooses his niche to be around his own type ($A_i \subset I_i$), he maximizes the probability of having retweet-able tweets in the following periods. And from lemma 2, we know that retweeting is optimal, i.e. there is a profitable deviation from the original MPE for all agents. Note that this is regardless of cost of following other users.

Additionally, from Lemma 3, we know that users find it profitable to follow more individuals than in the one-shot game. In fact, when $v > c$, the benefit of following the marginal agent is always greater than the marginal cost, therefore, users have incentive to follow all the agents up to $\theta_j = \theta_i \pm \omega$ whose niche intersection with i is just a point. The intersection between i's niche with any agent who is $\theta_j + \epsilon$ is the empty set (following agents beyond this point only decreases his own benefit without providing any retweet-able tweets for i). Therefore, the maximum number of agents that i would follow is $\overline{x} = 2\omega$ which maximizes a (equation 4.8). Although following others is costly, at the margin, there is always a strictly positive increase in the number of followers as a result of following the marginal agents.

Therefore, in symmetric equilibria, all the agents find it profitable to deviate from the static equilibrium and follow more agents. So each agent would follow all the agents that are $\pm\omega$ away from his type and also is followed by the very same group of individuals. In other words, the set of followings and followers of any agent is the same and is more than the number in the static equilibrium.

Similarly, in Lemma 3 we argued that agents have incentive to increase their niche because as their niche increase, the intersection of niches also increases, which in turn increases the probability of retweeting. However, if i increases his niche to $\omega_i > \frac{2\alpha}{\beta}$, even the agent whose type is at the centre of this niche will get lower expected utility from following i, which means that all the other agents who are following i also get lower expected benefit if $\omega_i > \frac{2\alpha}{\beta}$. As before, because of symmetry, the optimal niche size is $\omega \leq \frac{2\alpha}{\beta}$ (with equality when $c = 0$) for all agents.

Note that earlier, we argued that agents choose their niche $A_i \subset I_i = [\theta_i - \frac{2\alpha}{\beta}, \theta_i + \frac{2\alpha}{\beta}]$ to maximize the probability of having retweet-able tweets, and also since agents follows other to receive tweets closest to his type, therefore, his retweets would also be closest to his type and combined with $|A_i| \leq \frac{2\alpha}{\beta}$, $\theta_i \in A_i$ would maximize the probability of retweeting. When $c = 0$, agents' niche is $\frac{2\alpha}{\beta}$ and they follow $\pm\frac{2\alpha}{\beta}$ other agents around their own type, i.e. total number of everyone's followings is $\frac{4\alpha}{\beta}$. ■

In summary, although individuals follow more agents than in the static case, each agent *filters the informations* for their followers. Because of the niche size, agents who are close in the type space (i.e. maximum $\pm\frac{\alpha}{\beta}$ away from each other) receive positive utility from hearing each others' tweets and they give and receive additional beneficial information through retweets. However, when $v > c$, agents follow others further away and hear tweets farther away from their type, but they the only retweet ones that are in their niche. This results in a *Rat Race*, where agents follow others whose tweets they dislike, only for the small chance that they may say something once in a while that is retweet-able. In other words, individuals follow others not for the benefit from the tweets, but because of the fear of under providing information and losing followers when everyone else is following more agents.

Corollary **2.** *Since the set of followers and followings of any agent coincide, the links between individuals are bi-directional.*

Proposition **3.** *The stable equilibrium networks are either full, star or line networks, depending on the distance between agents in the type space.*

- $\forall\theta_i, \theta_j$ *such that* $|\theta_i - \theta_j| \leq \frac{2\alpha}{\beta}$*, the stable network is the full network regardless of the cost of link formation.*
- *However, if* i, j *are such that* $\frac{2\alpha}{\beta} < |\theta_i - \theta_j| \leq \frac{4\alpha}{\beta}$*, then there is no direct link between* i, j *and the only stable structure between such agents is a bi-directional star with the centre agent* k *such that* $\max\{|\theta_i - \theta_k|, |\theta_j - \theta_k|\} \leq \frac{2\alpha}{\beta}$.
- *Finally, if* i, j *are such that* $|\theta_i - \theta_j| > \frac{4\alpha}{\beta}$*, then there is a bi-directional path from* i *to* j*, such that the distance between any two adjacent individuals on this path is less than* $\frac{2\alpha}{\beta}$.

Proof. We already argued that the links between agents are bidirectional. And since the set of followers and followings for any individual coincide, the stable network is a completely connected network with individuals whose distances in the type space is less than or equal to $\frac{2\alpha}{\beta}$. Note that this is regardless of the cost of link formation.

If individuals are father away in the type space, i.e. $\frac{2\alpha}{\beta} < |\theta_i - \theta_j| \leq \frac{4\alpha}{\beta}$, they would not follow each other directly, however, in the type space there is at least an agent k between i and j that both i, j follow k and k follows them. In fact every agent follows (and is being followed) by agents that do not follow each other directly, and therefore, are centres of a star network.

Finally, if the distance between agents is greater than $\frac{4\alpha}{\beta}$, there are $l_1, l_2, ..., l_k$ agents such that $\max\{|\theta_i - \theta_{l_1}|, |\theta_{l_1} - \theta_{l_2}|, ..., |\theta_{l_{k-1}} - \theta_{l_k}|, |\theta_{l_k} - \theta_j|\} = \frac{2\alpha}{\beta}$, i.e. there is a bidirectional path of length $k+1$ between i, j that defines the line network. ■

Note that this result is different from the standard network equilibria where the structure of the network depends on the cost and benefit parameters [Bloch and Jackson (2007)]. Here, we see full networks not because of the high benefit from following other agents or because of low costs of link formation. But rather because of the fear of under providing information and being unfollowed by others. For example, in this scenario, star or circle are never stable when the agents' types are close enough.

In summary, we find that by moving from static setup to the dynamic environment, the set of Nash equilibria shrinks considerably, i.e. we do not see equilibria that agents tweet far from their own type. This findings sheds light on why individuals on Twitter develop an 'expertise', that is, a range of subjects on which they tweet which not only is limited to a generally relatively number of subjects, but also contains *their own* are of interest. Of course, another straightforward reason for why one's area of interest coincides with the content of one's tweets is that when information acquisition is costly (contrary to what we have assumed), the relative cost of obtaining information to pass on in a tweet is lower for information that one would have gathered anyway out of a consumption motive. Our analysis highlights, however, that this argument can be applied to one's followings, and the corresponding retweets, as well. If following someone is costly, then the benefits of gathering information from that individual's tweets are amplified by being able to retweet them. The latter requires that one's tweet niche overlaps with one's preferences and will, as a result, generally counterbalance any inceptive to tweet in 'popular' niches as opposed to an idiosyncratic niche close to one's own interests.

Furthermore, the set of followers and followings coincide and we see only bidirectional links between connected agents. In this environment, the only stable structures are full, star or line networks, depending on the distance between agents in the type space.

We do not observe bipartite or circle networks, also star is never stable when agents are close in the type space. These results hold irrespective of the cost of following other individuals.

5. Efficient Network from Planner's Perspective

In the one period game, there are two sources of inefficiencies. When individuals are choosing whom to follow, they only consider their own benefit from others' tweets, and not the positive utility gain from having followers for the other agents. At the same time, when agents are choosing their niche, they only care about maximizing the number of their followers and do not take into account the benefit that their followers receive from their tweets. To see this more formally, let B_{ij} denote the benefit that i receives from the tweets of individual j and consider a small increase in the width of j's niche. From equations (3.3) and (3.4) we have for types $\theta_i \notin A_j$,

$$\begin{aligned}\theta_i \notin A_j \quad \Rightarrow \quad B_{ij} &= E_{x_j \in [\overline{x}_j, \underline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] \\ &= -\beta\Big|\theta_i - x_{j,\text{center}}\Big| + \alpha \\ \Rightarrow \quad \frac{dB_{ij}}{d\omega_j} &= 0,\end{aligned}$$

whereas for $\theta_i \in A_j$, we have

$$\begin{aligned}\theta_i \in A_j \quad \Rightarrow \quad B_{ij} &= E_{x_j \in [\overline{x}_j, \underline{x}_j]}\Big[-\beta|\theta_i - x_j| + \alpha\Big] \\ &= -\frac{\beta}{\omega_j}\Big[(\theta_i - x_{j,\text{center}})^2 + \frac{1}{4}\omega_j^2\Big] + \alpha \\ \Rightarrow \quad \frac{dB_{ij}}{d\omega_j} &= \frac{\beta}{\omega_j^2}(x_{ij}^2 - \frac{\omega_i^2}{4})\end{aligned}$$

It is easy to see that when $2x_{ij} \leq \omega_j$ ($x_{ij} = |\theta_i - x_{j,\text{center}}|$), which is the case in equilibrium, then $\frac{dB_{ij}}{d\omega_j} \leq 0$. In other words, increasing the width of one's niche has a *negative* effect on the followers, while decreasing the niche has a positive effect as long as $\theta_i \in A_i$. When $\theta_i \notin A_j$, the change in the niche has no effect on the agent. Note that no agent would follow another individual if the distance between his type and the centre of their niche is greater than $\frac{\alpha}{\beta}$.

To characterize the extent of these inefficiencies, in this sub-section we contrast these results by the solutions to a planner's problem. Specifically, we now set up the problem from a planner's perspective and characterize the optimal solution assuming this planner chooses both the niche size as well as number of followings and followers to maximize the overall expected utility. We also focus on the symmetric equilibria. The overall expected utility in this network is:

$$\max_{\{x,\omega\}} \sum_{i=1}^{N} EU_i = \max_{\{x,\omega\}} \sum_{i=1}^{N} \Big(- \frac{\beta x^3}{12\omega} - \frac{\beta \omega x}{4} + \alpha x + vx - cx \Big)$$

$$(5.1) \qquad = \sum_{i=1}^{N} \max_{\{x,\omega\}} \Big(- \frac{\beta x^3}{12\omega} - \frac{\beta \omega x}{4} + \alpha x + vx - cx \Big)$$

Where x is the number of followings and followers of any agents in the symmetric equilibria. By taking the first order condition with respect to x, we have:

$$(5.2) \qquad \frac{\partial EU_i}{\partial x} = 0 \quad \Leftrightarrow \quad x^e = \sqrt{\frac{4(\alpha + v - c)\omega_i}{\beta} - \omega_i^2}$$

In other word, conditional on ω, planner always chooses $x^e > x^i$, this is because of considering the added benefit from having followers. On the other hand, the first order condition with respect to ω at the equilibrium x^e is:

$$\frac{\partial EU_i}{\partial \omega} = 0 \quad \Leftrightarrow \quad \frac{\beta x^{e3}}{3\omega^2} - \beta x^e = 0$$

$$(5.3) \qquad \Leftrightarrow \quad x^e = \sqrt{3}\omega^e$$

After substituting equation (5.3) in (5.2), we have:

$$(5.4) \qquad x^e = \frac{\sqrt{3}(\alpha + v - c)}{\beta} \qquad \omega^e = \frac{\alpha + v - c}{\beta}$$

Now it is easy to see, when for example $c = 0$, if $v > \alpha$, then $\omega^e > \omega^i$ and $x^e > x^i$, i.e. planner chooses wider niches for agents, but also chooses more followers for each agent as well.

In the dynamic game, the planner faces the following maximization:

$$(5.5) \qquad \max_{\{x,\omega\}} \sum_{i=1}^{N} EU_i = \sum_{i=1}^{N} \max_{\{x,\omega\}} \Big(\frac{4\omega}{x^2 + 4\omega(1-x)} \Big[- \frac{\beta x^3}{12\omega} - \frac{\beta \omega x}{4} + \alpha x \Big] + (v - c)x \Big)$$

The first order conditions of the above equation are of order 4, so we use mathematica to plot $\frac{dEU_i}{dx}$ when $\frac{dU}{d\omega} = 0$, to draw some conclusions. Comparing figure 1 and 2, we see that holding all other parameters constant, the optimal number of followers from planner's perspective in both static and dynamic game is a decreasing functions of c and an increasing function of v in both dynamic and one-shot game. (To Be completed)

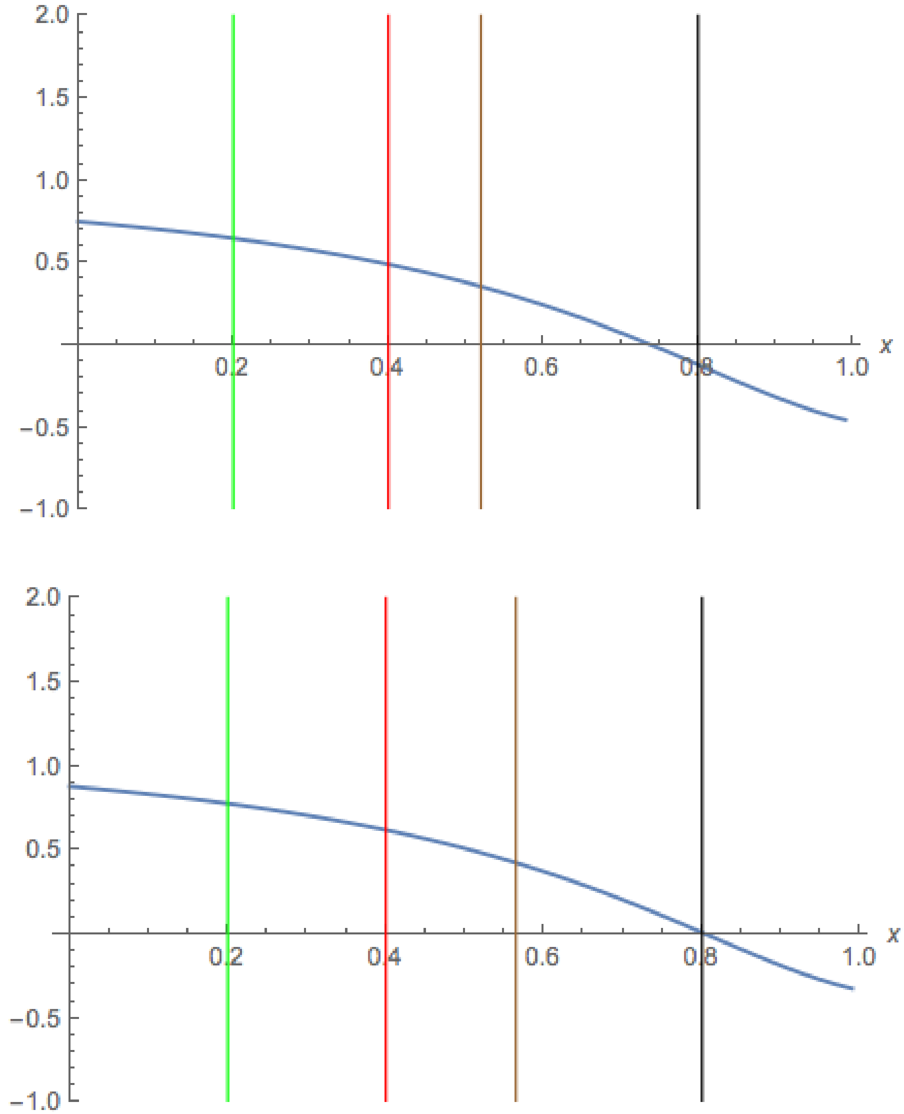

Figure 1. Top: $\beta = 5, \alpha = 1, c = 0.5, v = 1$
Bottom: $\beta = 5, \alpha = 1, c = 0.5, v = 1.13$
Red line is $\frac{2\alpha}{\beta}$ point, Green: Decentralized one-shot game solution, Brown: Planner's one-shot game solution, Blue: crosses zero at planner's dynamic game solution, Black line is $\frac{4\alpha}{\beta}$ point.
For middle ranges of c, number of followers in the one shot game is less than $\frac{2\alpha}{\beta}$, whereas planners chooses a number larger than $\frac{2\alpha}{\beta}$. Both planner's dynamic and static solutions are increasing with v.

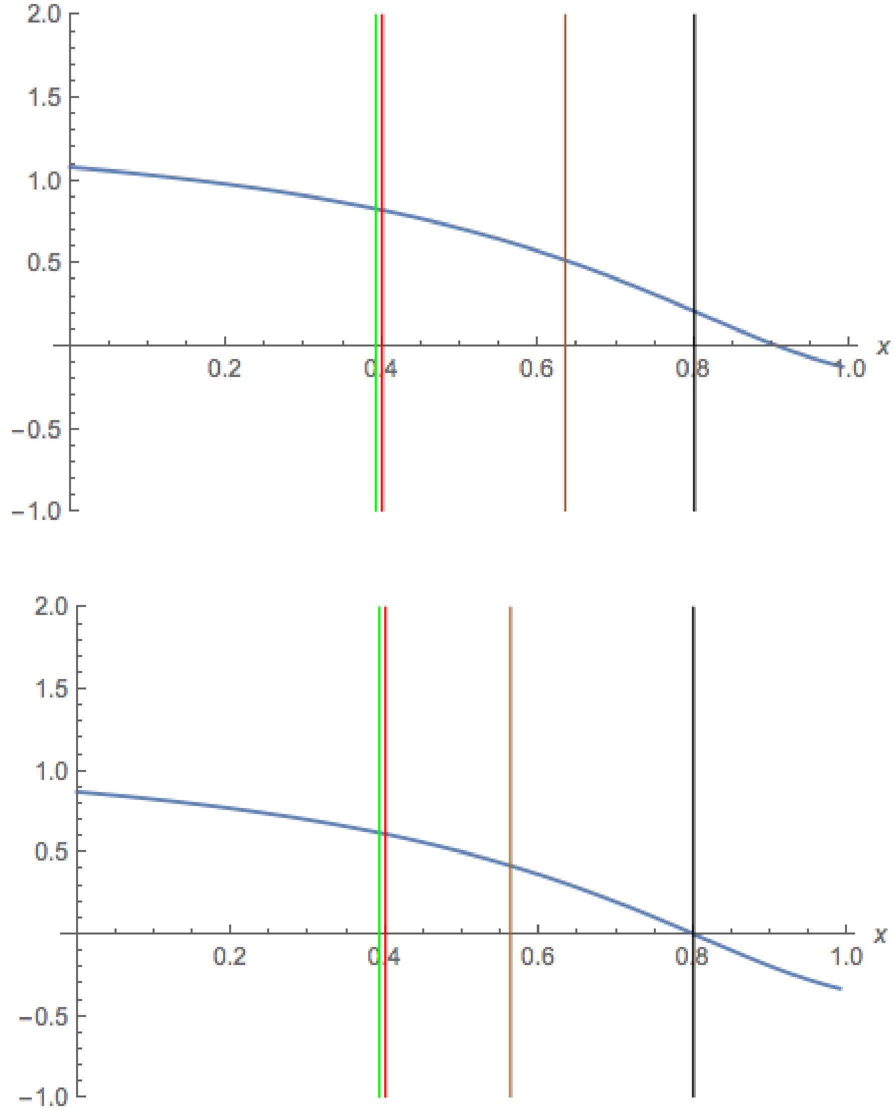

FIGURE 2. Top: $\beta = 5, \alpha = 1, c = 0.02, v = 0.85$
Bottom: $\beta = 5, \alpha = 1, c = 0.02, v = 0.62$
Red line is $\frac{2\alpha}{\beta}$ point, Green: Decentralized one-shot game solution, Brown: Planner's one-shot game solution, Blue: crosses zero at planner's dynamic game solution, Black line is $\frac{4\alpha}{\beta}$ point.
When $c = 0$, the number of agents followers is equal to $\frac{2\alpha}{\beta}$, but number of followers in the planners problem both in the one-shot game and the dynamic game depends on v: number of followers in both static and dynamic problem increases as v increase.

6. CONCLUSION

We formulated a theoretical framework to study the network configurations that arise in Twitter. We showed that when following other users is costly, retweets play an important role in creating shortest path in the network and as a result agents choose their niche to include their own type. We also showed that even when the cost of following is zero, there is an upper bound on the niche width as well as the number of agents that each user follows. Additionally, when $v > c$, agents follow others whose tweets they dislike, only for the small chance that they may say something once in a while that is retweet-able. In other words, individuals follow others not for the benefit from the tweets, but because of the fear of under providing information and losing followers when everyone else is following more agents. We also characterized the network configurations that would form in the Markov perfect equilibria: When users are close enough in the type space, the only stable network is a connected network with fully connected components, however, when agents are further away in the type space, star-like networks become stable. Circle network is never stable.

Finally, we contrasted our results with the planner's problem, and discussed two sources of inefficiency in the network. We show that the optimal number of followers from planners perspective is always greater than individuals' optimal followers in the one-shot game. However, in the dynamic game, depending on values of parameters in the model, planner's optimal number of followers maybe larger, equal or smaller than individuals' optimal choice.

References

[1] Bakshy, Eytan, et al. "The role of social networks in information diffusion." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.

[2] Bala, Venkatesh, and Sanjeev Goyal. "A noncooperative model of network formation." *Econometrica* 68.5 (2000): 1181-1229.

[3] Boyd, Danah, Scott Golder, and Gilad Lotan. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." *System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE*, 2010.

[4] Bloch, Francis, and Matthew O. Jackson. "The formation of networks with transfers among players." Journal of Economic Theory 133.1 (2007): 83-110.

[5] Conover, Michael, et al. "Political polarization on twitter." *ICWSM*. 2011.

[6] Goel, Sharad, Duncan J. Watts, and Daniel G. Goldstein. "The structure of online diffusion networks." *Proceedings of the 13th ACM conference on electronic commerce*. ACM, 2012.

[7] Golub, Benjamin, and Matthew O. Jackson. "How homophily affects the speed of learning and best-response dynamics." *The Quarterly Journal of Economics* 127.3 (2012): 1287-1338.

[8] Goyal, S. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2009.

[9] Halberstam, Yosh, and Brian Knight. "Are Social Media more Social than Media? Measuring Ideological Homophily and Segregation on Twitter." 2013.

[10] Hermida, Alfred, et al. "Share, like, recommend: Decoding the social media news consumer." *Journalism Studies* 13.5-6 (2012): 815-824.

[11] Hui, Cindy, et al. "Importance of ties in information diffusion." *Poster presentation at Workshop on Information In Networks* (WIN). 2010.

[12] Jackson, Matthew O. *Social and economic networks*. Princeton University Press, 2010.

[13] Jackson, Matthew O., and Asher Wolinsky. "A strategic model of social and economic networks." *Journal of economic theory* 71.1 (1996): 44-74.

[14] Jones, Martin, and Robert Sugden. "Positive confirmation bias in the acquisition of information." Theory and Decision 50.1 (2001): 59-99.

[15] Kawamoto, Tatsuro. "A stochastic model of tweet diffusion on the Twitter network." *Physica A: Statistical Mechanics and its Applications* 392.16 (2013): 3470-3475.

[16] Ko, J., et al. "Model for Twitter dynamics: Public attention and time series of tweeting." *Physica A: Statistical Mechanics and its Applications* 404 (2014): 142-149.

[17] Libai, Barak, Eitan Muller, and Renana Peres. "The social value of word-of-mouth programs: acceleration versus acquisition." *Marketing Science* WP_22_2009 (2009).

[18] Morales, A. J., J. C. Losada, and R. M. Benito. "Users structure and behavior on an online social network during a political protest." *Physica A: Statistical Mechanics and its Applications* 391.21 (2012): 5244-5253.

[19] Rui, Huaxia, Yizao Liu, and Andrew Whinston. "Whose and what chatter matters? The effect of tweets on movie sales." *Decision Support Systems* 55.4 (2013): 863-870.

[20] Situngkir, Hokky. "Spread of hoax in Social Media." (2011).

[21] Tarbush, Bassel, and Alexander Teytelboym. "Friending: a model of online social networks." (2013).

[22] Wang, Beidou, et al. “Whom to mention: expand the diffusion of tweets by @recommendation on micro-blogging systems.” *Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2013.

[23] atts, Alison. “A dynamic model of network formation.” *Networks and Groups*. Springer Berlin Heidelberg, 2003. 337-345.

[24] Wu, Shaomei, et al. “Who says what to whom on twitter.” *Proceedings of the 20th international conference on World wide web*. ACM, 2011.

[25] Yang, Jiang, and Scott Counts. “Comparing Information Diffusion Structure in Weblogs and Microblogs.” *ICWSM*. 2010.