Follow the Herd:

Exploring Information Aggregation in Networks with Application in Social Media

Fadi Egho

(California Polytechnic State University)

June 12, 2020

1 Introduction

Network Analysis has been enjoying a fair share of research amongst high-level Economists for decades, now. One might argue that it made the discrete jump from its mathematical cousin, graph theory, back when Morris H DeGroot published his seminal paper, "Reaching Consensus" (1974). DeGroot's paper, and the constant development of his model until today has marked a noticeable shift in how scientist think about and analyse complex systems. That being said, it is rather surprising and unwarranted how underexposed undergrad and Master level Economic students are to network theory and its wide array of applications. For that matter, this paper will serve as a soft introduction to Network Analysis for students unfamiliar with the subject, and will doubly serve as refresher for advanced readers who have not visited the topics in a while.

As a learning experience, the paper will introduce one of the classical works in sequential learning models by Banerjee, which closely examines herding behaviour (1992). To complement this theoretical model, the paper will introduce a simulation aspect to drive an intuitive approach to understanding the consequences of Banerjee's model. The paper will then transition by introducing the DeGroot model and discuss how it differs from Banerjee's. Finally, we'll examine some concepts from graph theory that mostly overlap with network analysis, as the paper culminates in analysing activity of so called ISIS fanboys on Twitter.

The rest of the paper proceeds as such: Section 2 will begin with an overview of current literature by introducing a simple herd model and providing a solid background in the DeGroot framework. Section 3 will expand on Banerjee's model in more depth, and discuss closely results and finding from the simulation, while Section 4 will switch gears to building a foundation in network analysis through real-world data. Finally, Section 5 concludes the analysis and discusses possible future extensions to this research.

2 Literature Review

Economic agents in analysis are often burdened with the rationality assumption and perfect, or at least symmetric, information. In a simple network structure, with a sequential learning framework, Banerjee (1992) breaks this mould by modelling the spread of information across players and their consequent actions based on such accumulative information (faulty or otherwise). Banerjee claims that this framework explains why we sometimes arrive at inefficient equilibria in various economic and social contexts, for example. Specifically, the model depicts how a rational individual chooses to disregard their own signal and follow what "everyone else is doing". Banerjee labels this as herd behaviour and considers the negative[1] externality it inflicts on society; i.e., subsequent players, too, ignore their private signals and follow the herd.

While somewhat simplified, Banerjee's model is nonetheless intuitive and informative. There are N players each with an identical, risk-neutral utility function, which is maximised by choosing between a set of assets indexed from [0, 1], such that $a(i)$ is the i^{th} asset, and $z(i)$ is the return[2]. Only one asset, $a(i^*)$, has excess returns and each player has a probability α of receiving a signal, s_n which tells her with probability β which i is i^*. Since Since decision rules will be discussed more closely in a later section, let us highlight Banerjee's main findings: (i) equilibria patterns are inefficient, over iterations of the game, and further, (ii) there is a non-zero lower bound to the probability of the entire population not choosing $a(i^*)$, and finally (iii) herd externality is self-perpetuating, which may help explain irrational outcomes like stock bubbles, or really any environment where an agent disregards a signal to follow external noise/information. There are certain tie-breaking conditions Banerjee assumed in order to arrive at his conclusion, however. These address how an agent would operate when he is indifferent between listening to a private signal and following another player's lead.

Golub and Sadler modified Banerjee's model to make it more realistic by varying the

[1] Note that Banerjee identifies this as a negative externality; as this paper will demonstrate, there are situations where herd behaviour is a positive externality, contingent on signal quality

[2] Details of this model will become relevant in later section of the paper

amount of information a player is able to observe (2017). In contrast, each player n makes a binary choice $x_n \in \{0,1\}$ about what $\theta \in \Theta$ might be, given a prior, $q_0 = \mathbb{P}(\theta = 1)$. The decision is mapped onto $u(x,\theta)$ such that

$$u(x=1,\theta=1) > u(0,1) \quad and \quad u(x=0,\theta=0) > u(1,0)$$

Instead of seeing every past decisions, however, the player n now observes $B(n) \subseteq \{1,2,3,\dots,n{-}1\}$, an arbitrary subset of players. Hence, with x_m being the action observed in the subset, the player's information can be denoted as, $I_n = \{s_n, x_m\}$[3]. Further, Golub and Sadler show the optimal strategy for player n, σ_n can be defined as:

$$\mathbb{E}_\sigma[u(\sigma_n,\theta)|I_n] \geqslant \mathbb{E}_\sigma[u(\sigma'_n,\theta)|I_n]$$

They then elegantly constructed an asymptotic outcome which summarises how information aggregates in accordance with the large-sample principle [5], provided that s_n is *i.i.d.*

$$\lim_{n\to\infty} \mathbb{E}_\sigma[u(x_n,\theta)] = q_0 u(1,1) + (1{-}q_0)u(0,0)$$

This also presents a counter, albeit not a complete contradiction, to Banerjee's results, where with a large enough sample, n makes the correct choice. This is not surprising, but to drive the point further, let us define the probability of making the wrong choice, $\varepsilon = \mathbb{P}(x_n \neq \theta)$; following the same logic, $\lim\limits_{n\to\infty} \varepsilon = 0$. In other words, n cannot make a mistake knowing everyone else's signal; she would "asymptotically" learn.

Although the two models discussed above are intuitive and useful to a point, a structural constraint is present; they lack a dynamic aspect which would otherwise allow us to study players' behaviour over time. As a natural extension, Golub and Sadler provide a concise introduction to the DeGroot Model, in the same paper [5]. The model utilises Markov

[3]Basically, information of player n is comprised of both private and public belief just as in Banerjee's framework; also in this case, $\alpha = 1$, so every player receives a signal

Chains[4] to map out paths of each player's belief at any given time, t. The versatility of the approach allows us to examine a network regardless of whether it is complete or not (interconnectedness of all agents). More specifically, model estimates a transition rule:

$$x_i(1) = \sum_j W_{ij} x_j(0) \tag{1}$$

where $x_i(1)$ is player i's belief at time $t = 1$, $\mathbf{W}$ is the $n \times n$ stochastic matrix reflecting each player's influence on one another, and $x_j(0)$ is the prior of player $j \in B(n)$ (so logically, the disconnect between player i and j would be represented by a zero at W_{ij}). Instantly, the intuitive consequence to explore here, is how the properties of a Markov matrix would translate onto the DeGroot model; i.e., what does an irreducible stochastic matrix (or analogously, a complete network[5]) imply? What is the connotation of a long-run steady-state in network analysis?

Not surprisingly, Golub and Jackson (2010) demonstrate that the convergence of a stochastic matrix (given irreducibility and aperiodicity) translates to what they call "long-run consensus" of beliefs in a network. To lay ground for this, let us generalise (1) into the entire network of players:

$$x(1) = \mathbf{W}x(0) \tag{2}$$

this is equivalent to updating prior (beliefs) of these players from period $t = 0$ to period $t = 1$. By properties of matrix powers, we can extend this indefinitely:

$$x(t) = \mathbf{W}\ \mathbf{W}\ \cdots \mathbf{W}\ \mathbf{W}\ x(0) = \mathbf{W}^t x(0) \tag{3}$$

[4]For a concise treatment of Finite Markov Chains, check out Thomas Sargent's lecture (2020)

[5]By definition of irreducibility, all states must communicate with one another, which in a network, translates to completeness. One cannot help but wonder if the idea of network completeness was derived from the definition of irreducibility, or whether it sprung independently. As I dearly value my time, I'll leave such enquiry to those interested in mathematical history

So by constructing $\mathbf{W}$ which reflects each player's prior and their influence on one another, we can essentially model the evolution of opinions over time in a Markov framework. More to Golub and Jackson's point, asymptotically, $\lim_{t\to\infty} \mathbf{W}^t x(0)$ converges to a *unique*, non-negative eigenvector that reflects the long-run consensus of the population[6]. Specifically, this "eigenvector of centrality" specifies the strength of influence of each player on the final consensus. This is a powerful notion, insofar that given a network, players with strong beliefs will influence those with weak beliefs, and over time, consensus will emerge as beliefs of those at the top softens, of those at the bottom strengthens, leading all players to meet somewhere in between. This result is guaranteed when both aperiodicity and irreducibility are satisfied. Golub and Jackson consider having "self-trust", or a player having some belief of her own, a sufficient condition for aperiodicity and treat this as the general case; nonetheless, they do show examples of where aperiodicity is satisfied without such assumption.

The applied nature of this paper will dictate the two following sections. Having explored both static and dynamic approaches to network analysis, this paper will attempt to implement methodologies discussed above to come up with two models that test theory boundaries. The first model will simulate Banerjee's game and contrasts its results with his findings [1]. The second model examines data from social media scraped by Fifth Tribe [9], and utilises techniques from the dynamic framework to extract possible insights from the data, all the while reporting any relevant limitations.

3 Simulated Model

3.1 Banerjee's Game

There are N utility-maximising players presented with a set of assets indexed from 0 to 1. As mentioned earlier, $u(a(i^*)) = z(i^*) > 0$ and $u(a(i)) = 0 \; \forall \; i \neq i^*$. We take α to represent

[6]Proof of uniqueness is outside the scope of this paper; for more details, see Chapter 8 in Meyer (2000) for discussion of Perron-Frobenius Theorem

signal *prevalence* in a population, and β to represent signal *quality*[7]. Additionally, decisions are made in sequence, and a player's prior on which $a(i) = a(i^*)$ is uniform, when he doesn't have a signal. That being said, there is an additional constraint imposed on the first player; to minimise first-mover distortions, player one's default pick is a(0) when he lacks a signal, otherwise he chooses according to signal. Further, Banerjee assumes successive players must abide by the following herd-minimising rules:

1. A player will follow his signal, i', if:
   (a) at least one previous player has picked $a(i')$
   (b) or, no other asset has been picked more than once (with the exception of $a(0)$)
2. If 1a is *not* satisfied, the player will pick the asset $a(i \neq 0)$ that is selected the most but at least twice; in case of a tie, he will pick asset with lower index value
3. If a player does not have a signal:
   (a) he will choose a(0) if no other a(i) has been chosen
   (b) he will pick the asset $a(i \neq 0)$ that is selected the most but at least twice
   (c) if neither 3a nor 3b is satisfied, he will pick a(i) with highest index value chosen

It's also crucial to note that the probability of picking $a(i^*)$ even with a bad signal is still non-negligible; at a failure rate of $1-\beta$, the probability of picking $a(i^*)$ follows a discrete uniform distribution, given the number of assets from which to choose. This consequently dictates the utility of a high (low) quality signal to be positively (negatively) correlated with the number of assets available. While this point was not particularly relevant in Banerjee's paper, a pivotal extension to his model that this paper will focus on, is how varying degrees of signal prevalence and quality affects the equilibrium outcomes.

Finally, after setting up the algorithm that replicates these rules above, five variations of α-β pairs were simulated for the total of 10,000 runs:

[7] Recall $\alpha = P(\text{player } n \text{ receives } s_n)$, and $\beta = P(s_n \textit{ signals } a(i^*) \textit{ correctly})$

Table 1: N = 40, assets = 21

Variation	α Value	β Value	# Simulations
High Prevalence\High Quality	.85	.85	2,000
High Prevalence\Low Quality	.85	.15	2,000
Medium Prevalence\Medium Quality	.5	.5	2,000
Low Prevalence\High Quality	.15	.85	2,000
Low Prevalence\Low Quality	.15	.15	2,000

To clarify, each simulation is a single game consisting of forty players sequentially abiding by the decision rules above. Along with the simulations, the following attributes were calculated for each run:

- *Influence*: an indicator denoting whether the game's first two movers had a signal
- *Optimal Rate*: percentage of players in a given game who picked $a(i^*)$, signal aside
- *Signal Prevalence*: percentage of players receiving a signal
- *Herding Index*: a compound variable calculating proportion of signal defiers in a game

It is important to note that signal is strictly private; no one player can "observe" any other player's signal. Any herding arises naturally out of probabilities and the rules of thumb highlighted earlier. As such, those rules were constructed to mimic what would happen in a simplified world. Furthermore, to motivate extensions in reality, the set-up of the simulations was geared to allow for as much granularity as possible; differentiating between influenced and uninfluenced outcomes was briefly discussed but left without treatment in Banerjee's paper. This, in addition to examining outcomes with respect to varying signal quality and prevalence will be the main highlights of the analysis.

3.2 Findings

Table 2 summarises key results of the simulation. Outcome of *Optimal Rate* across different signal types are higher when the first or second player in a game has a signal, particularly when the quality of the signal is high. This is captured by 10 percentage point difference in the High-High and Low-High categories. Examining Herding Index gives a slightly different perspective, on the other hand. Herding Index, as defined above, is the proportion of agents who received a signal, but followed another player, nonetheless. The results indicate that having an "Influencer" does not reflect changes in outcome when α is high.

Table 2: Summary Results of Optimal Choice & Herding

Variation (Prevalence-Quality)	Optimal Rate		Herding Index	
	Influenced	Not Influenced	Influenced	Not Influenced
High-High	0.868	0.774	0.169	0.166
High-Low	0.231	0.218	0.691	0.668
Med-Med	0.469	0.430	0.581	0.542
Low-High	0.607	0.507	0.265	0.184
Low-Low	0.150	0.109	0.803	0.660

On the other hand, with low signal prevalence, having an influencer is associated with higher rates of herding. Although one might argue when α is low, statistics of Herding levels will be prone to volatile outcomes by pure design. Since low signal prevalence essentially implies lower numbers from which to calculate proportions. While this justification makes sense, and the author at first glance thought so, upon closer inspection, this proved to be inaccurate. Running multiple simulations with different random seeds proved robust to simulation volatility. More importantly, the differential trend persisted *consistently* between Influenced and non-Influenced sub-populations[8]. One interpretation of this lies in $1-\beta$, the probability of choosing the wrong signal. With a scarce signal, a first mover who happen to receive a distorting signal, will inadvertently lead on an uninformed crowd. So even when a few others end up receiving a correct signal, it is already too late to consider the quality of private signal, as the trend has already been set. This is supported by the variance

[8]While meta-analysis is outside the scope of this paper, check out figure 4 in the appendix for some strikingly consistent results using different random seeds. After running different combination of simulations, volatility on the population level was not present

between low and higher quality signal. When the β is low, the difference in Influenced vs. Uninfluenced Herding is more prominent.

Additionally, as centrality measures do not give the whole picture, it is important to examine the entire population of data. Figure 1 depicts the distribution of *Optimal Rate* across different signal quality and prevalence. Densities are stacked by *Influence*, and are contrasted by distribution of *Signal Prevalence*, which is represented by the black line. The figure shows a much more nuanced story when it comes to medium-prevalence, medium-quality signal. There is a distinct bifurcation in decision. This is where relying on an average point estimates would have given zero information about what is going on, since the majority of trials are concentrated at the extreme tails. Clearly, average signal did not result in average outcome here, since most of the runs either arrive at a pareto-optimal equilibrium or a severely inefficient one. The instance here corroborates Banerjee's second conclusion about volatile equilibria, but only conditional on signal quality; High-Prevalence, High-quality conditions are quite stable in variations. As tautological as it may sound, the takeaway here is: ambiguity triggers volatility.

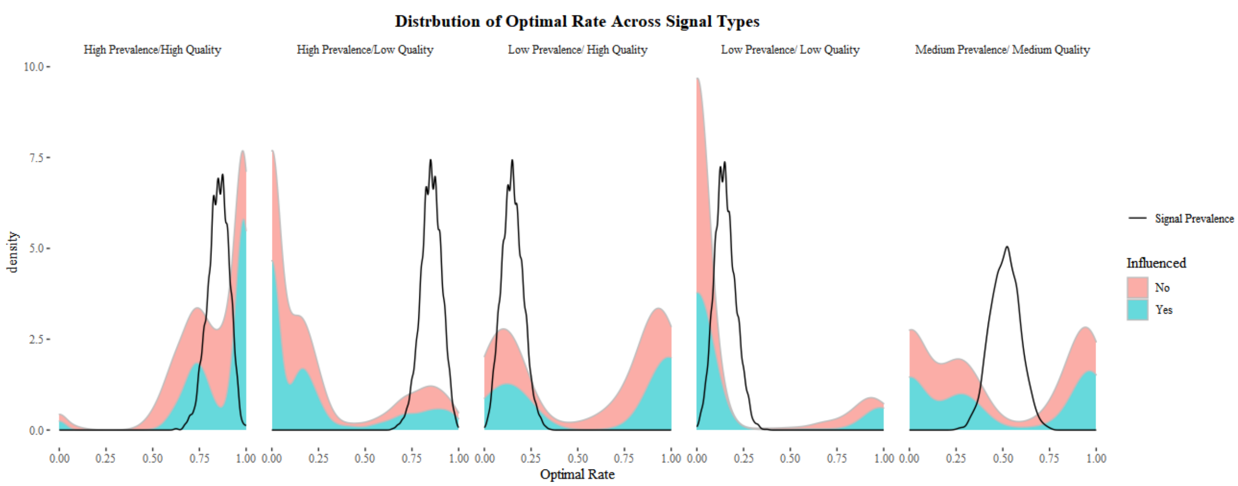


Figure 1: Distributional Characteristics

We conclude this section with the Table 3 reflecting two-tailed t-test results of optimal rates.

The test fails to reject the null hypothesis in the high prevalence low quality sample.

Table 3: Two-sided t-test on Influence vs uninfluenced optimal rates

Variation	statistic	t_{df}	p-value	Lower CI	Upper CI
High-High	-3.139	50.96	0.0028***	-0.154	-0.033
High-Low	-0.302	57.19	0.763	-0.098	0.072
Med-Med	-1.87	804.2	0.061*	-0.080	0.002
Low-High	-4.81	907.	1.73e-06***	-0.1409	-0.059
Low-Low	-2.55	822	0.011**	-0.074	-0.01

Note: * p < 0.1; ** p <0.05; *** p <0.01

Otherwise, in the four other categories, we reject that average optimal rate does not vary conditional being *Influenced*.

4 Empirical Model

4.1 Twitter's Terrorist Network

This section will delve into a more realistic framework and take into consideration concepts Golub & Sadler [5], and Golub & Jackson [4] discussed with respect to the DeGroot learning model. To develop a basic learning framework, this paper sources its dataset from Kaggle [9]. The data were published by Fifth Tribe, a digital agency that serves non-profits and governmental agencies. In 2016, Fifth Tribe scraped tweets generated by ISIS fanatics all around the world. These tweets spanned from late 2015 to mid 2016 (roughly corresponding to Islamic State's peak notoriety period). The data captures variables such as *username*, count of $followers$, *Description* of profile, *time stamp*, and the *tweets*, themselves. The data points do not provide explicit information on which user follows which, nor is there a way to obtain this information publicly; in fact, majority of users accounted for in the data have been suspended by Twitter. To build a network, therefore, this paper relies on a seminal aspect of a tweet: *mentions*.

Before getting into further detail about the dataset, however, let us introduce some fundamental nomenclature in graph theory. Many of the term have already been introduced here, but carry different, more generalised labels in maths. As a reference, this paper relied on the manual for the *igraph* package in R, which provides a rich introduction to network structures [2]. Briefly speaking, a network is comprised of nodes and edges. A node can be a person, a website, an address, or any entity that sends and/or receives a signal of some sort. Edges, on the other hand, are the channels through which those nodes can relay signals[9]. Furthermore, a network can be directed (follower/followed) or undirected (mutual). A network can be mapped onto an adjacency matrix, which again, is a generalised form of the stochastic matrix that was introduced earlier along with the DeGroot Model. A key distinction is in implementation: a stochastic matrix is used to describe transition states, usually with respect to a process convergence [8], while an adjacency matrix need not follow the same parameter restrictions, and could simply be used to reflect linkages[10].

Additionally, to summarise a network, three key statistics are typically used: Degree Centrality, Closeness Centrality, and Betweenness Centrality. Degree is the number of edges a node has; this typically denotes how well connected a node is. Closeness is the mean distance from one node to another, which can typically imply better access to information, in the social network context. In contrast to Banerjee's model, Closeness would be an alternative to modelling signal *quality*; i.e., it would be a function of closeness, as opposed to being modelled deterministically. Finally, Betweenness measure the degree to which a node lies on the path between other nodes. In a social network, this can take on the meaning of an information broker, or a gatekeeper to a cluster of tightly connected nodes. Again, as compared to Banerjee's model, Betweenness would internalise signal prevalence into a network model, rendering it a function of betweenness.

Reflecting back on the dataset itself, it is critical to highlight how the network was constructed for this analysis. As mentioned earlier, there is not an organic way to link nodes (users) to one another, since the no information, except the count, was provided about each user's followers. Therefore, each tweet was parsed for mentions, and edges were constructed

[9] just with node, a signal can take on many different meanings depending on the context

[10] In other words, a Markov matrix is a very special type of an adjacency matrix

based on the mentioned user and the "tweeter". The intuition behind using mentions as a proxy for a signal/edge is that it may help uncover insights on how this population interacts, and by utilising the frequency at which each user mentions another, it may give us some guidance on who the leader(s) of the herd are.

Upon parsing the tweets, observations that lacked mentions were consequently dropped. To proceed from here, two tables had to be derived for the *igraph* package to build the network. The first is a list of unique players preferably along with a numeric identifier. For the second, after some tricky data manipulation, mentions that each user tweeted were turned into counts, and each pair of sender-recipient were aggregated to reflect their "lifetime" interaction. With this information a network is constructed. At this point, it is important to to distinguish another statistic by which to describe the network. Just as not all nodes are equal, neither are the edges linking them. A user who is followed by 1,000 other users sending a signal carries more weight than another user doing it with other only 50 followers. Likewise, a mention carries more weight when there is a longer history of interactions between a pair of sender-recipient, than otherwise. This concept of weighting edges based one the users they link, gives rise to Strength [2]. Strength is a weighted degree centrality and assigns more gravity towards some users than others. in the case of this data, after deriving the number of time a user mentioned each recipient over time, it was appropriate to construct the following weight:

$$weight_{edge} = ln(r_{followers_i} * mentions_{ij}) \tag{4}$$

where $r_{followers_i}$ is a value reflecting where user i ranks in number of followers (more followers, the higher value, r), and $mentions_{ij}$ is the total number of times user i has mentioned user j. It's important to keep in mind, since the framework is directed, $mentions_{ij} \neq mentions_{ji}$ normally. The reason why it's reasonable to take the product of the two to prevent the model from overestimating the weight of edges from low-count users from who bombards shout-out daily. Likewise we take the rank, instead of the actual number of followers, as to not skew the score too heavily for users who have orders of magnitude more followers. Finally, the log transformation is a final adjustment as to prevent the model from exploding, in case we have the worst of both scenarios.

4.2 Findings

While the layout of how to approach the analysis was fairly straightforward, managing the data processing stage was fairly tasking. Even after getting the data in the appropriate format for *igraph* to transform into a network, the adjacency matrix turned out to be quite sparse, despite the sizeable amount of data points. Before delving into the details, however, let's first address the basics.

Degree centrality is simply calculated by summing over the rows of the adjacency matrix, since those values reflect which nodes user i is linked to. Moreover, an un-weighted adjacency matrix will simply reflect ones and zeros for each row. Otherwise, you'd just re-weight edges according to, for example, (4). Alternatively, if one wishes to construct a stochastic matrix as in the DeGroot model, one would have to re-weight probabilities accordingly[11]. The Softmax formula elegantly makes such transformation [7]:

$$\phi_{softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

As mentioned earlier, one problem the dataset suffered from was that the adjacency matrix was sparse, and the network was *not* strongly connected. This violated the sufficient conditions for (3), and because of this, there is no direct path for signal to pass through every node of the network; hence, there is not a unique eigenvector that would quantify a given user's influence over the long run information aggregation (no consensus). This result is not completely discouraging, as an eigenvector of centrality still exists, and although it does not

[11] uniform priors wouldn't make sense, if there are already weights assigned to each edge

describe the general equilibrium, it still gives us some information about the degree to which the user influences the local neighbourhood, $B(n)$ that he or she observes.

Another important measure is betweenness, which, according to the *igraph*, is calculated using the simple algorithm:

$$\sum_{i \neq k} \frac{n_{ik}^{(j)}}{n_{ik}}$$

where $n_{ik}^{(j)}$ is the number of shortest possible paths between i and k that go through j, and the denominator is the number of shortest possible paths from i to k, in general. As shown in Figure 2 below, there are clear opportunities for signal throttling or signal rebranding by users floating at the top of the plot. This is marked by their extreme Betweenness score, as compared to the other nodes. Basically, any time a user would tweet something, it is bound to go through *MaghrabiArabi*, for example, before reaching anyone else in the network. Curiously enough, upon glancing over a few of his tweets, his profile description proclaims that he is just an "Analyst watching events in the ME", and his first tweet reads, "SubhanAllah [Glory be to god] their are mujahideen with the black flag". This speaks to the external validity of betweenness and how it can tie back to this paper's treatment of Banerjee's model. More specifically, the effect of herding as a function of Signal Prevalence. When a node is positioned to be the gatekeeper to other members of the network,that node can modulate the prevalence of a signal (quality aside) to serve a particular agenda.

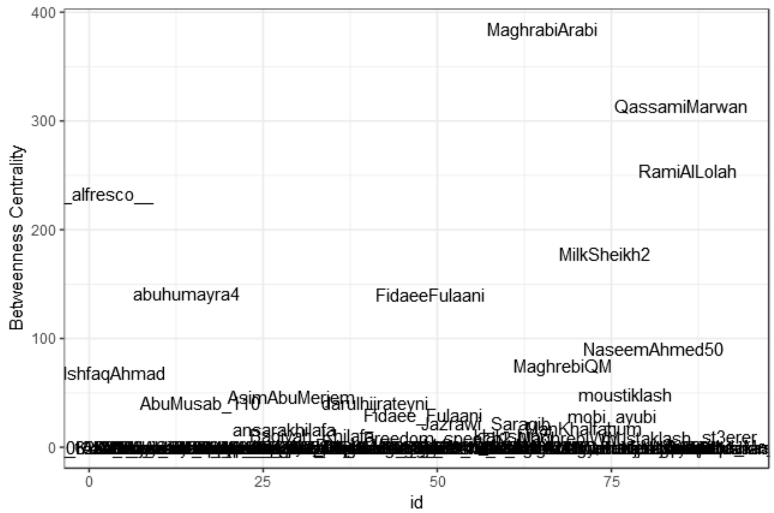


Figure 2: Gatekeepers of The Twitter Mob

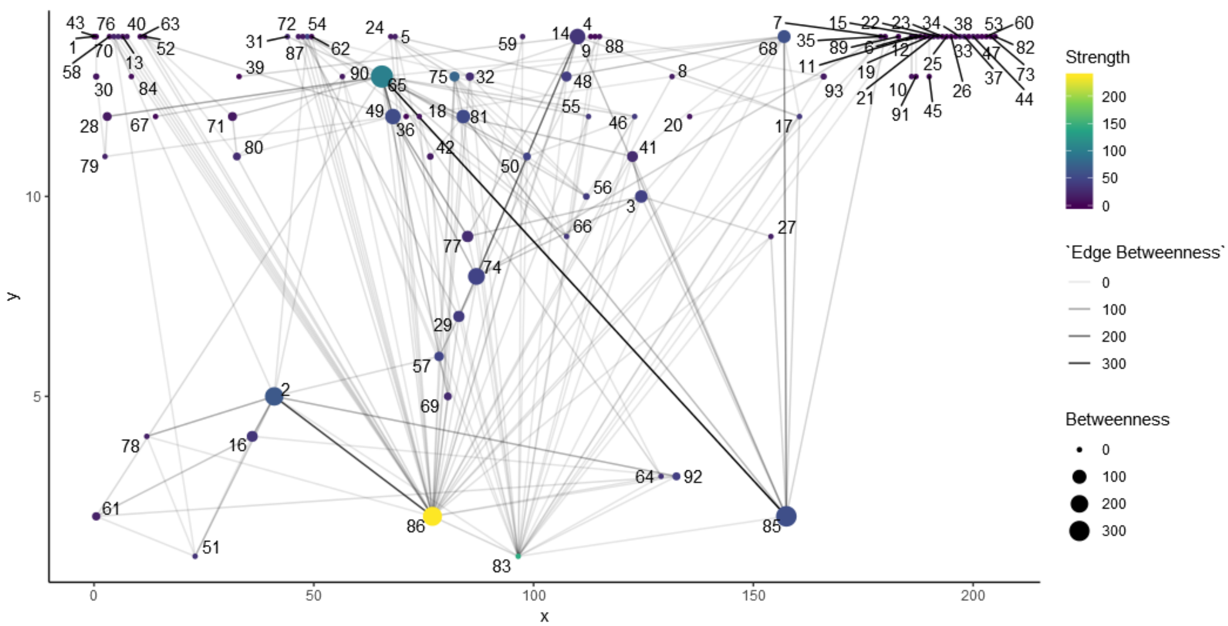
Table 4 displays some the key player's attributes, including the eigenvector. Also note that the *Degree* measure has been normalised by dividing it by the number of nodes. As for Closeness centrality, the measure is unfortunately invalid, also because of incompleteness of the network.

Table 4: Key statistics of the network

id	user	Degree Centrality	Betweenness	Eigenvector	Followers
67	MaghrabiArabi	0.098	384	0.468	158-353
87	QassamiMarwan	0.054	314	0.305	1,600
88	RamiAlLolah	0.196	254	1	24K
2	__alfresco__	0.058	233	0.359	560
76	MilkSheikh2	0.040	177	0.181	1,000-1,300
51	FidaeeFulaani	0.049	140	0.232	380

That has been the biggest obstacle the dataset has presented. Since missing followers cannot be recovered to complete the network, the only other alternative is to downsize the network enough where all nodes are connected; this renders the data too small to be viable for any statistical analysis. Nonetheless, exploring the Twitter data was a great introduction into Network analysis and hopefully encouraged readers who were unfamiliar with the subject to think much more about it, and ideally start exploring the topics discussed here independently. This section will end by exploring one last figure that summarises our network rather succinctly. Figure 3 maps out three attributes about the Twitter network. Size of the node reflects the betweenness score discussed earlier, while the thickness of the edges identifies important channels of communication, and lastly, the brighter nodes reflect powerful influencers. A useful by-product of this plot, is that it also clusters smaller neighbourhoods of users together. This helps the the viewer get a better understanding of the overall network structure.

Figure 3: Sugiyama-style Interpretation of the network



5 Conclusion

This paper has introduced some powerful concepts pertaining to herding behaviour, sequential learning models, and applied network analysis. Although the Twitter dataset was not perfect, it nevertheless served its purpose as a unique opportunity to developing genuine interest in Network Analysis and Graph Theory. Possible extensions to this paper include, but are not limited to, meta-analysis into some of the simulations adopted in this paper.

More specifically, an opportunity lies in closely examining trade-offs of relaxing some of the assumption of the simulated game. Furthermore, as Golub and Sadler pointed out (2019), the notion of a non-existent feedback loop would be quite apt to model echo chamber behaviour, or persuasion bias. Such framework may provide great insight about the behaviour of certain extremist groups, like the ones examined on Twitter.

References

[1] Abhijit V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992. http://www.jstor.org/stable/2118364.

[2] Gábor Csárdi and Tamás Nepusz. *Network Analysis and Visualization*, 2006. https://cran.r-project.org/web/packages/igraph/igraph.pdf.

[3] Morris H. Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480137.

[4] Benjamin Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, February 2010. https://www.aeaweb.org/articles?id=10.1257/mic.2.1.112.

[5] Benjamin Golub and Evan Sadler. Learning in social networks. *Available at SSRN 2919146*, 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2919146.

[6] Carl Meyer. *Matrix Analysis and Applied Linear Algebra.* Society for Industrial and Applied Mathematics, 2000. http://www.cse.zju.edu.cn/eclass/attachments/2015-10/01-1446085870-145420.pdf.

[7] Sebastian Raschka. What is softmax regression and how is it related to logistic regression, 2016. http://kdnuggests.com/2016/07/softmax-regression-related-logistic-regression.html.

[8] Thomas Sargent and John Stachurski. Finite markov chains, 2020. https://python.quantecon.org/finite_markov.html.

[9] Fifth Tribe. How isis uses twitter, 05 2016. https://www.kaggle.com/fifthtribe/how-isis-uses-twitter.

A Appendix

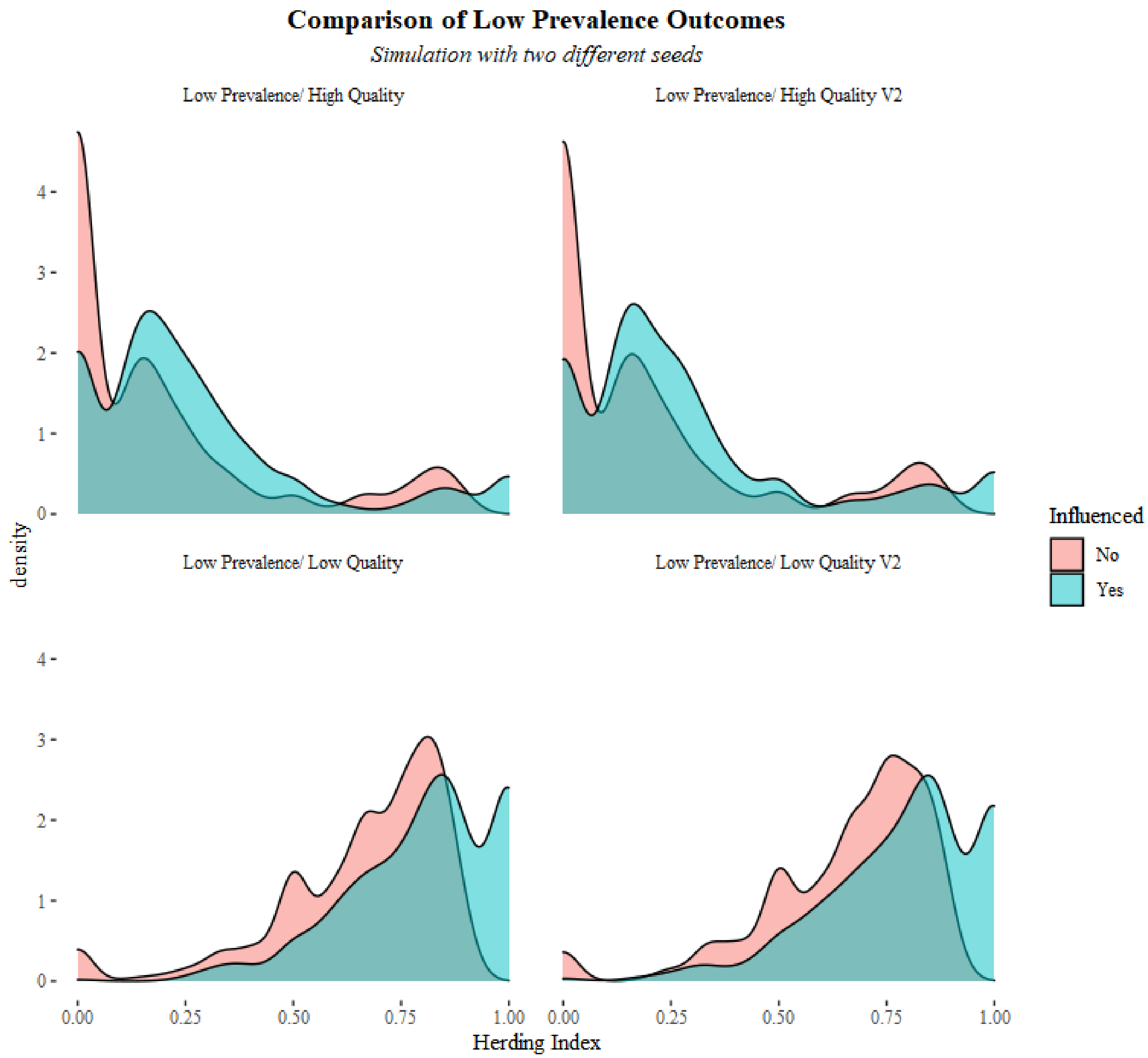


Figure 4