GRATIA: COMPUTING SOCIAL CAPITAL AS ENGAGEMENT AND BELIEF REVISION

Gaurav Koley

Master of Technology Thesis

June 2019

ज्ञानमुत्तमम्

International Institute of Information Technology, Bangalore

GRATIA: COMPUTING SOCIAL CAPITAL AS ENGAGEMENT AND BELIEF REVISION

Submitted to International Institute of Information Technology, Bangalore
in Partial Fulfillment of
the Requirements for the Award of
Master of Technology

by

Gaurav Koley
IMT2014019

International Institute of Information Technology, Bangalore
June 2019

Thesis Certificate

This is to certify that the thesis titled **Gratia: Computing Social Capital as Engagement and Belief Revision** submitted to the International Institute of Information Technology, Bangalore, for the award of the degree of **Master of Technology** is a bona fide record of the research work done by **Gaurav Koley**, **IMT2014019**, under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Srinath Srinivasa

Bengaluru,

The 25th of June, 2019.

GRATIA: COMPUTING SOCIAL CAPITAL AS ENGAGEMENT AND BELIEF REVISION

Abstract

Social capital, or the goodwill that one accrues by virtue of their skills, actions and values, play an important role in building sustainable communities and organizations. Most online social spaces have some kind of a reputation scoring system for participants in the online space. However, a dearth of computational foundations for social capital has resulted in reputation management systems susceptible to hyperinflation, mob dynamics, and invisibility due to non-use. Studies suggest that using engagement as a factor indicates influence and reputation better.

In this research, we develop a computational model for social capital, based on two factors: engagement and belief revision. Engagement refers to the impact or traction arising from a user's actions in the community, and belief revision refers to the way such impact influences others' beliefs. Engagement is measured in terms of any action that is scarce in the community. Belief revision is modeled using DeGroot learning. A diffusing computation model is used to develop the concept of an Authority Rank and a Citizen Rank for a participant, for a given topic. This model also gives us two parameters λ and β which characterize the community and the effect of engagement.

The proposed system is tested on NetLogo simulations as well as on an academic pre-print management system, where users get to read pre-prints of technical papers, in an online reader. Mechanisms are designed to estimate the amount of sustained attention received by resources uploaded on this portal, and propagation of this attention on to its

creators.

Experiments show that the with the Authority Rank (AR), it is possible identify agents in a community that enjoy a highly engaged and entrenched audience which existing models cannot.

Acknowledgements

I would like to begin by expressing my sincere gratitude and appreciation to my thesis advisor **Prof. Srinath Srinivasa**. Over the time, I have worked with him he has patiently guided me and helped in refining my flashes of insights into substantial ideas. I have enjoyed the long hours of countless discussions with him on various topics.

I would like to thank Jayati Deshmukh for her invaluable contributions that shaped this thesis into its current form. Her work on Netlogo and the simulations built for this thesis helped in understanding the nature of this work. A part of this thesis utilises the Netlogo models she built.

I want to thank Anshumaan Agrawal for the numerous heated discussions I had with him about the nature of engagement and attention and what why do people engage with others. His inputs and suggestions have been extremely useful.

I have been a part of the Web Science Lab (WSL) since 2017 and it has been a one of kind association. The research talks, lab meetings and general discussions at WSL have greatly influenced me. Being a member of the WSL, I got the opportunity to do some amazing work in a team and gained skills like time management and team work apart from the technical knowledge.

I would like to thank the members of my thesis defence committee, **Prof. G Srinivasaraghavan** and **Prof. Manish Gupta** for providing their valuable feedback and suggestions leading to an improved version of the thesis.

Special thanks to Vishesh Ruparelia for helping me build the Gratia portal which forms an important pillar of this thesis. I also want to thank Satvik Ramaprasad and the entire CircuitVerse committee for proving anonymized user data from the CircuitVerse portal without which this thesis would have taken much longer to complete.

Special thanks to all my friends here at IIIT-B who have always stood by me and have been my family away from home.

Finally, I would like to thank my parents for supporting me in all my endeavours to the hilt having faith in me and guiding me at every step in life.

— Gaurav Koley

Contents

Abstract iii

Acknowledgements v

List of Figures x

List of Tables xii

List of Abbreviations xiii

1 Introduction 1

2 Social Capital 3

2.1 What is Social Capital 3

2.2 Definitions of Social Capital 3

2.2.1 Individualistic Perspective on Social Capital or Linkages 4

2.2.2 Groupist Perspective on Social Capital or Bonds and Bridges . . 5

2.2.3 Capability based approach to Social Capital 6

2.3 Effects of Social Capital 6
2.3.1 Positive Effects 6
2.3.2 Negative Effects 7
2.4 Existing Algorithms for Measuring Influence 8

3 Model **9**

3.1 Social Capital as Engagement 9
3.2 Strokes and Engagement 9
3.2.1 Engagement 11
3.2.2 How to get more engagement 12
3.3 Network Characteristics 13
3.3.1 DeGroot Belief Revision 14
3.3.2 Authority Rank 15
3.3.3 Citizen Rank 18

4 Experiments **22**

4.1 Gratia 22
4.2 CircuitVerse 24
4.3 Netlogo 25

5 Results and Comparisons **27**

5.1 Authority Ranks 27

5.2 Citizen Ranks 29

5.3 Comparisons 32

5.3.1 Comparing across Networks 32

5.3.2 Comparing with PageRank 33

6 Conclusions **36**

6.1 Applications 37

6.2 Future Work 37

Bibliography **39**

List of Figures

FC3.1 v gets stroked by $u_1 \dots u_4$ . . . 12

FC3.2 Sample Actor Attention Network . . . 13

FC3.3 With increasing λ, the relative attention is skewed to give more weight to high attention . . . 17

FC3.4 With increasing λ, AR values change and stabilize after a while. . . . 18

FC4.1 Actors create resources which are then consumed by other Actors . . . 23

FC5.1 Authority Ranks for Gratia sorted in decreasing order. Stable λ = 51 . 27

FC5.2 Authority Ranks for CircuitVerse sorted in decreasing order. Stable λ = 35 . . . 28

FC5.3 Authority Ranks for a Barabasi–Albert network sorted in decreasing order. Stable λ = 12 . . . 28

FC5.4 Authority Ranks for a Random network sorted in decreasing order. Stable λ = 26 . . . 29

FC5.5 Authority Ranks for Watts–Strogatz network sorted in decreasing order. Stable λ = 4 . . . 29

FC5.6 Citizen Ranks for Gratia sorted in decreasing order. Stable $\beta = 46$ . . 30

FC5.7 Citizen Ranks for CircuitVerse sorted in decreasing order. Stable $\beta = 42$ 30

FC5.8 Citizen Ranks for a Barabasi–Albert network sorted in decreasing order. Stable $\beta = 12$ 31

FC5.9 Citizen Ranks for a Random network sorted in decreasing order. Stable $\beta = 11$ 31

FC5.10 Citizen Ranks for a Watts–Strogatz network sorted in decreasing order. Stable $\beta = 14$ 32

FC5.11 PageRank and Authority Ranks for Gratia users with $\lambda = 10$ 33

FC5.12 PageRank and Authority Ranks for Gratia users with $\lambda = 51$ 34

FC5.13 PageRank and Authority Ranks for CircuitVerse users $\lambda = 10$ 34

FC5.14 PageRank and Authority Ranks for CircuitVerse users $\lambda = 35$ 35

List of Tables

TC5.1 Comparing λ across networks 32

List of Abbreviations

AR .... Authority Rank

CR .... Citizen Rank

IIITB .... International Institute of Information Technology Bangalore

CHAPTER 1

INTRODUCTION

Social Networking Sites(SNS), such as Facebook, Twitter, Instagram, Whatsapp pervade our lives. These services have connected people far and wide. To many, the online social networks are much more real than the physical social networks that they live in.

Online social networks and the spaces they create influence us in a variety of ways. We are constantly bombarded with information by advertisers and influencers who analyze and mine these social networks using what is known as Social influence analysis. By analyzing the flows of influence among users based on social networking site data, we can:

1. identify nodes of high influence;
2. understand people's behavior in a social context;
3. provide a theoretical basis for public policy and decision making.

Social influence analysis in social networks, therefore, has important applications and social significance.

However, social networks have undergone massive expansion in both scale and volume and that has brought numerous challenges in measuring social influence for a given user. This is compounded by the lack of a mathematical definition for measuring social capital and influence. Further, differing views on what constitute as the major factors for modeling social influence lead to several models with no way to effectively integrate these disparate factors for measuring influence. The factors being used as proxies for influence, namely: Likes, Upvotes, Claps etc are based on impulsive actions and are susceptible to hyperinflation, mob dynamics, and invisibility due to non-use.

Also, it is hard to properly characterize the relationship between social network structure and social influence and the causality of its effect. [1]

Studies [2, 3] suggest that using engagement as a factor indicates influence and reputation better. These studies also show that social influence is highly dependent on the online platform and that each platform is experienced in its own way.

Current methods of measuring social influence to identify influential individuals/organisations in a social network involves, either:

- using the reputation management system in place which focus on gamefied impulsive actions which are superficial in nature and very rarely reflect substantial influence; or
- using network centrality measures, link topological ranking measures and completely ignoring the quality of the links that point to the nodes in the network.

In this work, we address this issue by building computational models for measuring engagement and propose two new influence metrics: Authority Rank and Citizen Rank based on engagement and belief revision.

CHAPTER 2

SOCIAL CAPITAL

2.1 What is Social Capital

The concept of social capital has evolved over the years and has numerous definitions, each pertaining to a different domain. Sociologists, psychologists, network scientists, economists and others define social capital in their own way and it's difficult to come up with a singular definition that everyone will agree. However, broadly social capital refers to the factors that makes social groups function effectively namely, a shared sense of identity, trust, cooperation, shared norms, shared value system and reciprocity.

2.2 Definitions of Social Capital

Bourdieu [4] defined social capital as "the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition". Bourdieu calls the benefits that one receives by being a member of a particular family, community, class or institution as 'potential resources'. These resources are experienced in the exchange for 'trust' and 'gratitude' [4].

Bourdieu's definition has been criticized by many [5, 6] because of its focus on relationships built on complex social interactions based on memberships and access to social circles. He calls social capital as a 'collectively-owned capital' for a network's members, but provides analytical methods of measuring social capital based on the individual's position in the network structure. This makes practical applications of his model difficult.

Robert Putnam [7, 8] defines social capital as "features of a society that help facilitate and coordinate actions within that society. These features include social networks, norms of reciprocity, and levels of trust". Putnam draws upon Bourdieu's definition and combines it with the concept of 'civic culture' as presented by de Tocqueville [5, 9].

Since, both social networks and civic behaviour are characterized through people's memberships to social circles and associations, Putnam's approach has been criticized for presenting a model where civic culture, economy and social capital are all inextricably linked together [5, 6]. Further, Putnam ignores trust and reciprocity from the definition of social capital, something that Bourdieu emphasised [5, 6]. Putnam, like Bourdieu, also tends to ignore importance of gender and ethnicity in creating social capital [9] and that it can be used for promoting anti-social activities [6].

Several scholars proposed the use of social network analysis to study the social capital of groups and individuals; to emphasis the value of relationships and networks to maintain social capital. [10, 11]

2.2.1 Individualistic Perspective on Social Capital or Linkages

The individualist approach explores the idea that an actor's network or their position within the network affects their social capital. Individuals benefit from the following structural features of their networks:

1. size of an actor's ego network, where small networks with strong ties provide material and emotional support and large networks with weak ties provide access to information and resources [12, 13];
2. actor's position within the network, where having a broker position leads to being able access a greater diversity of resources than others [11, 14].

For example, Burt [11,14] based his empirical work on social capital on 'brokerage', where an actor acts as a broker or mediator between two other actors who are not linked directly.

Other empirical work on social capital from an individualist perspective include a 'resource generator' by Van Der Gaag, et al, [13] where actors were asked which relations (eg. acquaintances, family and friends) provided access to different kinds of resources. The 'position generator' by Nan Lin [15, 16] is another such empirical work which locates the hierarchical position of various people whom the actor is connected to.

2.2.2 Groupist Perspective on Social Capital or Bonds and Bridges

The groupist perspective focuses on a network's structural features and how the rise and maintenance of reciprocity and trust is enabled by that network structure. Coleman [17] posits that trust and the feeling of mutual obligation is higher among members of a complete network where all the actors are connected to each other.

Empirical research has shown social capital to correlate with increased perceived credibility of the flow of information within a network and people trusting each other on a long-term basis [11]. In addition, Granovetter [18] notes that this closed network structure persuades friends to behave honestly amongst themselves.

2.2.3 Capability based approach to Social Capital

In recent times, researchers have looked at Sen's Capability Approach [19, 20] to explain certain properties exhibited by social capital. Sen calls social capital as an endowment, i.e. "a set of means to achieve a life people reason to value".

Migheli [21] projects the concept of social capital onto a capability approach based framework. Here social capital acts as a catalyst for the achieving new capabilities and functionings. Migheli bases his model on the concept and causative effects of relational ontology.

Bertin and Sirven [22] adopt a rights-based definition of social capital and show that with Sen's capability approach it is possible to analyze the social interactions linked to the access to social resources. This approach treats social capital as an asset made up of the informal social rights that an agent can acquire through their social network. This distinguishes social capital from the social network it is embedded in and its norms and values; thereby giving a more precise evaluation of people's capabilities. However, this approach is restricted to the methodological difficulties in empirically evaluating people's capabilities.

2.3 Effects of Social Capital

2.3.1 Positive Effects

The potential positive effects of social capital can be seen through the influence of social connections. Friends and family helps individuals in lots of ways – economically, socially and emotionally.

Social capital is valuable for getting higher on the social ladder, especially for finding employment. Several economists including Granovetter [18] emphasise the role of close friends and family, as well as casual acquaintances, in finding jobs. Burt [14] focuses on the lack of close ties as a motivator for knowledge sharing and mobility of individuals. He identifies information and influence as the benefits derived from social networks. Another benefit may be understood as a form of social support involving co-operation and reciprocity not based on any immediate paybacks but improving the collective well-being of the society.

Mwangi and Ouma [23] show that social capital increases the network reach of an individual and thereby enhances financial inclusion through increased access to informal loans in Kenya. In their words, "...the higher the number of groups one pledges loyalty to, the higher the probability of accessing a loan."

Studies have shown that social capital encourages social trust and membership, reduces health risks amongst children and adolescents [24] and discourages individuals from engaging in harmful activities like smoking and binge drinking. [25]

Social capital is also linked with greater well-being according to self-reported survey measures; [26] and reduced crime [27].

Social capital also helps businesses. In Bowling Alone, [8] Putnam observes that the formal and informal cooperation between the startup companies in the Silicon Valley has led to their collective success. Humphrey and Schmitz [28] highlight how "trust-based relations between economic agents have been seen as part of the competitive advantage of manufacturing enterprises in Germany, Japan and parts of Italy ..."

2.3.2 Negative Effects

Highly entrenched social capital can hinder people as well. Close knit communities usually have strong social bonds, with the individuals relying heavily on relatives and others of the same community for support. The lack of social bridges that can connect them to the wider society, can also turn them into outsiders and hinder their social development and economic upliftment. [26]

Social capital can also be put to harmful use as well. The trust and reciprocity that allows mafia, criminal gangs and cults to operate is also a form of social capital. [26]

2.4 Existing Algorithms for Measuring Influence

The current methods of evaluating social influence include link topological ranking measures, centrality measures etc. [1]

Link topological ranking These measures are computed from the links that connect nodes in a network. E.g.:

- **HITS** This ranking measure gives two scores to a node in a network, a hub score, which estimates the value of the connections it has and a authority score, which estimates the value it holds or generates.
- **PageRank** This measure computes a probability distribution which represents the likelihood of a random walk across the network of landing at a particular node in the network.

Centrality These measures are computed from the network structures. E.g.:

- **Degree Centrality** which is defined as the number of links (both incoming and outgoing) that a node has;
- **Closeness Centrality** which is defined as the average length of the shortest path between a node and all other nodes in the network; and
- **Betweenness Centrality** which is defined as the number of times a node exists in the shortest paths between two other nodes in the network.

CHAPTER 3

MODEL

3.1 Social Capital as Engagement

Social Capital can be defined as a function of positive engagement. When actors engage with each other positively on issues of mutual interests or tags they are likely to establish a shared identity which would result in development of trust and familiarity. This encourages sharing information, sharing personal experiences with others, endorsing positive behaviour or discouraging negative behaviours, all of which are important for sustaining the community.

3.2 Strokes and Engagement

Any social interaction is composed of strokes. Berne [29] defined a stroke as the “fundamental unit of social action.” [30] A stroke is an act implying recognition of another actor’s presence either explicitly or non explicitly. Thus, any social interaction involving a stroke will have two roles of actors: a stroke initiator and a stroke receiver…For an actor u engaging with a resource created by another actor v would mean a stroke between u and v. Here u recognises the presence of v and is hence said to be ”stroking” v. Thus, u is the stroke initiator or consumer and v is the stroke receiver or creator.

Therefore, any social interaction will have two roles of actors:

Creator creates resources on the platform and influence other actors. Each creator has a topic vector g aligning to which they create content.

Consumer consumes resources on the platform and get influenced. Each consumer has an interest vector s and they consume content which closely matches their interest vector.

Then level of engagement from actor u with an interest s_u to an actor v with a topic vector g_v is:

$$\varepsilon_{u\to v} \propto \text{similarity}(g_v, s_u)$$

where

$$\text{similarity}(g_v, s_u) = 1 - \mid \text{difference}(g_v, s_u) \mid$$
$$\text{difference}(g_v, s_u) \mapsto [-1, 1]$$

We therefore define a stroke:

$$s = \langle\, u,\ v,\ t,\ \varepsilon_{u\to v}\,\rangle$$

where

$$u = \text{stroke initiator}$$
$$v = \text{stroke receiver}$$
$$t = \text{time of stroke}$$
$$\varepsilon_{u\to v} = \text{amount of engagement in the stroke}$$

In different scenarios, the act of engagement or stroke refers to different kind of social actions, e.g., reading a former research paper in bibliography networks and subsequently citing it, or responding to a tweet on Twitter. The action represents a directional influence from v to u, therefore, the action performer u and v are called consumer and creator, respectively. We consider there is an attention flow from the consumer to the creator.

A Social Stream is the sequence of strokes generated by the users from a beginning time to the current time t, which is denoted by $S_t = \{s_1, s_2, ..., s_m\}$. [31]

From a social stream, it is possible to generate a Social Network like the one that we will see in section 3.3.

3.2.1 Engagement

The model is heavily dependent on the way engagement is measured. Not all measures of engagement work well with this model. Our definition of measurable engagement is any action that has the following properties:

Scarce An agent should have only a finite quantity of that action. The finite quantity must also be small enough to be realistically be scarce. For example, Twitter allows a maximum of 2400 tweets in a day per user account. This number is finite and small enough to be realistically be considered as scarce. If the limit had been 240,000 tweets, although finite, it could not be considered as scarce since any user would hardly be able to reach the tweet limit.

Renewable The available quantity of the action that the actor can undertake should be renewed after some fixed duration. Continuing the example of Twitter, at midnight, the user can again post as many as 2400 tweets and they would be counted against that day's limit. Therefore, each day the limit counter is reset.

There are several actions which have the above properties and can be used as a measure of engagement. One of the simplest is time. The amount of time an actor spends engaging with another actor through their tweets, posts, etc can be tracked and used as a measure of engagement since the amount of time an agent can spend is limited in a day and is also renewed the next day.

Another measure can be retweets in twitter as a user can only post a maximum of 2400 tweets including retweets and this limit is reset everyday.

3.2.2 How to get more engagement

Since, $\varepsilon_{u \to v}$ depends on the similarity between g_v and s_u, the closer these vectors are, the higher the engagement will be.

Creators who want to increase their engagement can do so by tweaking their topic vector g to closely match the interests vector s of the actors that stroke them (i.e. consumers).

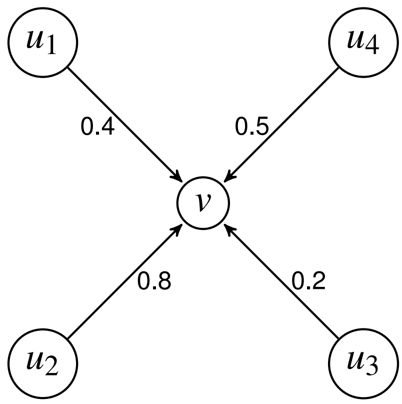

Figure FC3.1: v gets stroked by $u_1 \ldots u_4$

In Figure FC3.1, v receives strokes from $u_1 \ldots u_4$. The edge numbers refer to the ε values those strokes. To increase engagement, v would want to reduce difference(g_v, s_u) for all its neighbors. This can be done iteratively by updating its topic vector g though

Equation Eqn 3.1.

$$g_v = g_v - \sigma \cdot \sum_u \text{difference}(g_v, s_u) \quad \forall\, u \ni u \rightarrow v \qquad \text{(Eqn 3.1)}$$

where σ = adaptability constant

3.3 Network Characteristics

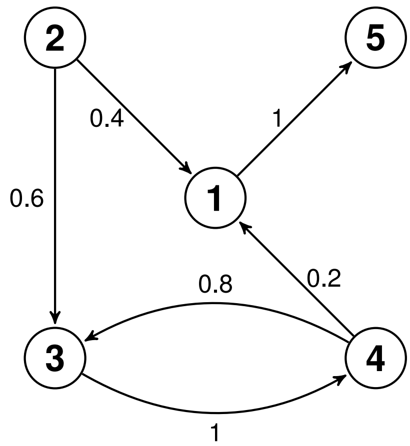

Figure FC3.2: Sample Actor Attention Network

A Social Network is a graph that reveals the attention pathways between the users on an online society, e.g., Twitter, DBLP, etc.. A social graph can be formulated as (V, E), denoted as N, where V is the vertex set and E is the edge set. Each edge $e \in E$ represents a nonempty set of strokes. Figure FC3.2 is an example of a social network where the edges in the graph are weighted by the attention paid.

Thus, in a social graph (V, E), an edge $e_{u \rightarrow v} \in E$ from vertex u to vertex v will have weight:

$$w(u, v) = \sum \varepsilon_{u \rightarrow v}\ \forall \text{ strokes between u and v} \qquad \text{(Eqn 3.2)}$$

We develop a bottom-up model of social capital, based on the engagement of a participant through their contributions. A decay function models the importance of recency. Since attention and thereby social capital is transitive, a diffusing computation model based on DeGroot's belief revision is used to develop the concept of Authority Rank (AR) and Citizen Rank (CR) for a participant.

3.3.1 DeGroot Belief Revision

To understand DeGroot belief revision, let us take a social network of n agents where everybody has an opinion on a topic, represented by a probability vector $p(0) = (p_1(0),\ldots,p_n(0))$. Actors receive no new information to change their opinions but can interact and communicate with their neighbors. Each link between actors (who knows whom) and their influence is represented as a belief matrix T where $T[i][j]$ measures the influence actor j exerts on actor i. T, thus, gives a weighted, directed graph where an edge $(i \rightarrow j)$ exists when $T[i][j] > 0$. The trust matrix T is stochastic, i.e. each rows is composed of nonnegative real numbers that sum up to 1.

Formally, in each time period agents beliefs are updated as:

$$p(t) = T \times p(t-1) \qquad \text{(Eqn 3.3)}$$

Unravelling the recursion, the t^{th} period opinions can be computed by

$$p(t) = T^t \times p(0) \qquad \text{(Eqn 3.4)}$$

When the network represented by T is fully connected, aperiodic and irreducible, the network's belief converge at:

$$P = \lim_{t \to \infty} T^t \times p(0) \qquad \text{(Eqn 3.5)}$$

We use this concept build Authority Rank and Citizen Ranks as the network's belief about the influence and affinity to the network respectively.

3.3.2 Authority Rank

Authority Rank is a measure of the impact of an actor in the social network. This can be interpreted as the network's belief about an actor's ability to impact it. For example, AR(5) = 0.09 can be understood as the network's belief that actor 5 can impact 9% of the network.

We construct the trust matrix T from the engagement strokes as defined by Eqn 3.2.

$$T[i][j] = w(i,j) = \sum \varepsilon_{u \to v} \ \forall \text{ strokes between u and v} \qquad \text{(Eqn 3.6)}$$

We say that an actor v's influence on actor u depends on the relative slice of attention paid by u to v. For example, let is consider Actor A who spends 4 hours engaging equally with actor C and 3 other actors on a social networking site and actor B who engages with only actor C for 1 hour. Although, the time (engagement measure in this scenario) spent by A and B on C is the same, A will be less influenced by C as compared to B. This can be explained as the following:

B engages on the social networking site solely to interact with C and, hence, if anything is posted by C or C moves away from the platform, B will be affected more than A.

Thus, relative engagement can be represented as:

$$r[i][j] = \frac{T[i][j]}{\sum_k T[i][k]} \qquad \text{(Eqn 3.7)}$$

where $r[i][j] \mapsto [0,1]$

We want to reduce the influence of low engagement and dampen effect the extremely high engagement as well. This is due to the fact that the law of diminishing returns holds true for engagement as well. Someone who is already highly engaged, any higher engagement doesn't lead to a linear increase in influence. Therefore, we use a softmax function to dampen lower engagement values and skew the engagement to give more weight to higher engagement values. Eqn 3.8 does this.

$$A[i][j] = \begin{cases} \frac{\exp(\lambda \cdot r[i][j])}{\sum_{k \neq i} \exp(\lambda \cdot r[i][k])} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \qquad \text{(Eqn 3.8)}$$

where $A[i][j] \mapsto [0,1]$
and λ = scaling parameter that represents point of valuable engagement

Here λ is a scaling parameter that is representative of the level of engagement required to judge another actor's worth as an authority. λ therefore depends on the measure of engagement being used as well as the network structure. Higher the median value of engagement for a kind of engagement measure, higher should be the λ. For example, if time is being used a measure of engagement, for a social network of videos, average time of engagement will be in 10s of minutes while for a twitter like network it will be lesser Therefore, λ will be smaller for a twitter like network than for a video network.

Figure FC3.3 shows the effect of λ. Higher the λ, more emphasis is given to higher engagement and consequently, low engagement is damped.

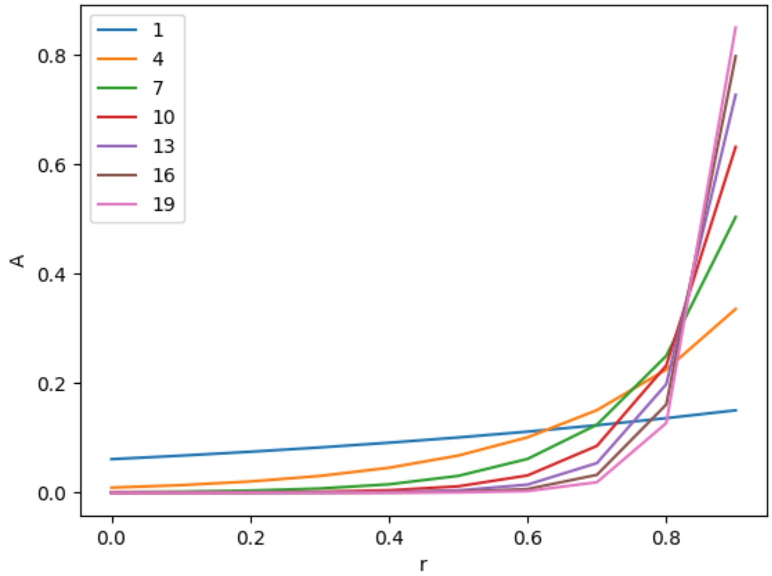

Figure FC3.3: With increasing λ, the relative attention is skewed to give more weight to high attention

Further $A[i][j]$ is strongly connected, aperiodic and irreducible, therefore, a convergence point for the network's belief exists.

$$A^* = \lim_{i \to \infty} A^i \qquad \text{(Eqn 3.9)}$$

and the left diagonal of A^* gives us the Authority Rank (AR) for the social network as:

$$AR(i) = A^* \times \vec{1} \qquad \text{(Eqn 3.10)}$$

Here, A is a non-Ergodic system, meaning the initial state of A determines the stable point of the system at the end of Eqn 3.9. Therefore, by varying λ, different results could be found for AR. Figure FC3.4 shows the change in the "volume" of AR with change in λ. To find the ideal λ, we search the space $[1, \infty)$ and settle for a value of λ for which subsequent increase in value doesn't change the AR values. The exact algorithm is described in Algorithm 1

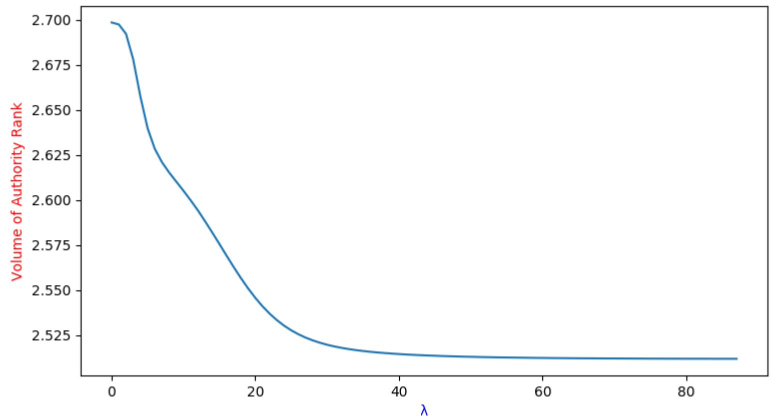

Figure FC3.4: With increasing λ, AR values change and stabilize after a while.

We pick the λ for which Authority Ranks stabilize. This stable value of λ changes across different social networks and different measures of engagement. For a network, the stable point of λ value indicates whether people prefer engaging with others on a deeper level or not. A small λ would represent that actor's prefer number of connections over deep engagement. A large stable value of λ would mean that actor's prefer engaging deeply than merely interacting with a lot of actors.

3.3.3 Citizen Rank

We have a similar process to generate Citizen Ranks. Citizen Rank is a measure of the perceived value of an actor's participation in the social network. This can be interpreted as the network's belief about the actor's contributions to preserving and improving the network. For example, CR(5) = 0.09 can be understood as the network's belief that actor 5 can is a stakeholder of the network with their stake being worth 9%.

We start with the transpose of the T matrix as defined in Eqn 3.6. With that we define our participation matrix as:

$$p[i][j] = \frac{T[j][i]}{\sum_k T[k][i]} \tag{Eqn 3.11}$$

Here $p[i][j]$ is the participation or engagement i has received from j. We run it through a similar scaling treatment to give disproportionately higher weightage to high engagement through:

$$C[i][j] = \begin{cases} \frac{\exp(\beta \cdot p[i][j])}{\sum_{k \neq i} \exp(\beta \cdot p[i][k])} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{Eqn 3.12}$$

where $C[i][j] \mapsto [0,1]$
and β = scaling parameter that represents point of valuable participation through engagement

Here β is a scaling parameter that controls the point at which the creator believes that the reader is engaged. β therefore depends on the measure of engagement being used, the kind of content on the network and the network structure. Similar to λ, a higher median value of engagement should reflect a higher β. Thus, for a social network of videos, β would be smaller for a twitter like network than for a video network.

Similar to $A[i][j]$, $C[i][j]$ is also strongly connected, aperiodic and irreducible, therefore, a convergence point for the network's belief exists at:

$$C^* = \lim_{i \to \infty} C^i \tag{Eqn 3.13}$$

and the left diagonal of C^* gives us the Citizen Rank (CR) for the social network as:

$$CR(i) = C^* \times \vec{1} \tag{Eqn 3.14}$$

Here, C, like A, describes a non-Ergodic system. Therefore, we search the space $[1,\infty)$ and settle for a value of β for which subsequent increase in value doesn't change the CR values. This stable value of β changes across different social networks and different measures of engagement. For a network, the stable point of β value indicates whether for the community the level of engagement is valuable compared to the number of interactions happening. A small β would represent that more interactions are better for network maintenance over deep engagement. A large stable value of β would mean that rich meaningful, highly engaging interactions are preferred.

The complete algorithm for computing Authority Rank and Citizen Rank from engagement values along with the ideal λ and β values is described in Algorithm 1.

Algorithm 1: Computing AR and CR using Engagement

```
let T be a |V| × |V| matrix of engagement.
T[u][v] = ∑ ε_{u→v} ∀ strokes between u and v; u, v ∈ V
A = T, C = T^⊤
for i ← 1 to |V| do
    for j ← 1 to |V| do
        A[i][j] = A[i][j] / ∑_j A[i][j]
        C[i][j] = C[i][j] / ∑_j C[i][j]
    end
end
AR_old = [], CR_old = []
for λ ← 1 to ∞ do
    for i ← 1 to |V| do
        for j ← 1 to |V| do
            if i == j then
                A[i][j] = 0
            else
                A[i][j] = exp(λ·A[i][j]) / ∑_{k≠i} exp(λ·A[i][k])
            end
        end
    end
    A* = lim_{i→∞} A^i
    AR = diagonal(A*)
    if AR = AR_old then
        break
    else
        AR_old = AR
    end
end
for β ← 1 to ∞ do
    for i ← 1 to |V| do
        for j ← 1 to |V| do
            if i == j then
                C[i][j] = 0
            else
                C[i][j] = exp(β·C[i][j]) / ∑_{k≠i} exp(A[i][k])
            end
        end
    end
    C* = lim_{i→∞} C^i
    CR = diagonal(C*)
    if CR = CR_old then
        break
    else
        CR_old = CR
    end
end
```

CHAPTER 4

EXPERIMENTS

Our measure of engagement is based on the amount of time a consumer spends on a creator either reading, viewing or responding to their works. We design several experiments to capture user engagement in online social networks and use this data to show the value derived from using the AR and CR metrics.

4.1 Gratia

Due to lack to actionable data from social networking sites, we built an academic pre-print management system, where users get to read pre-prints of technical papers, in an online reader. We capture the amount of time spent by the users viewing each paper, which gives us an estimate of the sustained attention and hence, the engagement received by said paper. Mechanisms are designed to attribute the attention received by papers to its creators.

The portal revolves around resources. Resources on the portal refer to PDF documents but the concept can be extended to video, audio and other types of content as well. Each resource has one or more "creators" who are the owner of that resource and any engagement with that resource will be attributed to them. For example, Figure FC4.1 has a resource "Resource 1" which has two creators: Actor A2 and Actor

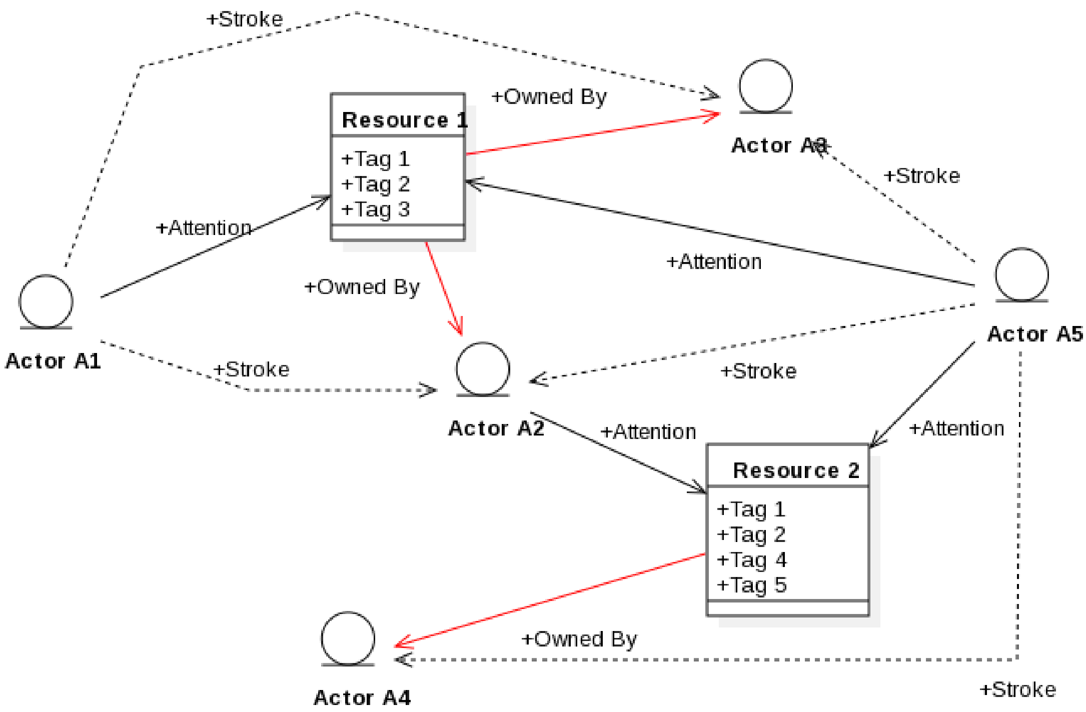

Figure FC4.1: Actors create resources which are then consumed by other Actors

A3. Each resource may have one or more tags associated with it.

When users view resources, the portal captures the amount of time users spend on engaging with a resource. This amount of time is then attributed to the "creators" of that resource by organising them as strokes as defined in Section 3.2. In Figure FC4.1, Actor A1 pays attention to Resource 1 and this is captured as the amount of time being spent. This time (k) is then attributed to Actor A2 and A3 separately as strokes. For example, if Actor A1 paid k amount of time to Resource 1 at timestamp t, then the engagement attributed to A2 and A3 will be as the following strokes:

$$s_1 = \langle\, A1,\ A2,\ t,\ k/2\,\rangle \text{ and } s_2 = \langle\, A1,\ A3,\ t,\ k/2\,\rangle$$

From a collection of such strokes, we can construct a trust matrix T and compute Authority Rank and Citizen Rank as specified in Sections 3.3.2 and 3.3.3.

The time spent on engaging with a resource is tracked on the server side with restrictions in place to ensure that a particular user can have only one resource open at a time. It is also made sure that the amount of time being counted towards engagement is only when the user is actively interacting with the open resource. Time tracking starts with the user's first interaction with a resource's webpage and ends when the user moves away from the page either by closing the browser or navigating to a different page. If a resource page is left open and hasn't been interacted with for 5 continuous minutes, the timer stops on the server side. This prevents the system from being gamed by empty engagement attacks.

For the experiments, the data was collected from the activities of 68 users. A total of 61 users spent 1534 minutes of time, collectively, engaging with 15 users across 23 resources leading to a graph made up of 253 strokes.

4.2 CircuitVerse

CircuitVerse is an easy to use digital logic circuit simulator which provides a platform to create, share and learn digital circuits. The platform has been designed for use by students, professionals and hobbyists alike for self-learning digital logic design. Apart from the simulator, users can create, learn, collaborate and share their work. CircuitVerse is currently used by several universities worldwide since it provides features for teachers to create groups and host assignments on the platform.

For our data collection procedure, the same backend and data model as Gratia was used for CircuitVerse as well. For the experiments, data was collected over a sample of 711 users. This led to 1540 strokes across 554 resources. A total of 343 users spent 6928 minutes engaging with 456 users.

4.3 Netlogo

To augment the data collected from Gratia and CircuitVerse, a series of simulations were made in Netlogo. Netlogo is a programmable multi-agent modeling environment used for designing multi-agent and network simulations.

Three kinds of directed networks of agents were simulated:

Barabasi–Albert Such networks are random scale-free networks generated using a preferential attachment mechanism. These networks contains a few nodes having unusually high degree in comparison to other nodes. Many natural and human-made networks like the Internet, citation networks are approximately Barabasi–Albert networks.

Watts–Strogatz This type of network follows small world properties, i.e. a network where most nodes are not neighbors of each other but most nodes can be reached from any node by a small number of hops or steps. This model closely models social networks.

Random This type of network is generated by randomly connecting two nodes with a predefined probability till the requirements of edges is complete.

All the above types of networks were simulated with 50 agents in each simulation. Each agent (called turtle in Netlogo) was given a topic value g and an interest value s. The model as described in Section 3.2 specifies vectors for t and s. However, for simplicity, we assume the network to be centered around a single tag or topic and hence a vector of length 1, i.e. a single value is enough to represent t and s.

Each agent then tries to align it's t value with the s values of it's readers (neighbouring agents which have a directed edge to said agent) using Eqn 3.1. Once the topic values stabilize across the network, agent's spend time engaging with each other. This time is determined by:

$$t_{u \to v} = 10 \times |g_v - s_u|$$

The AR and CR values and the λ and β values are then computed and observed for the network as specified by Algorithm 1.

CHAPTER 5

RESULTS AND COMPARISONS

5.1 Authority Ranks

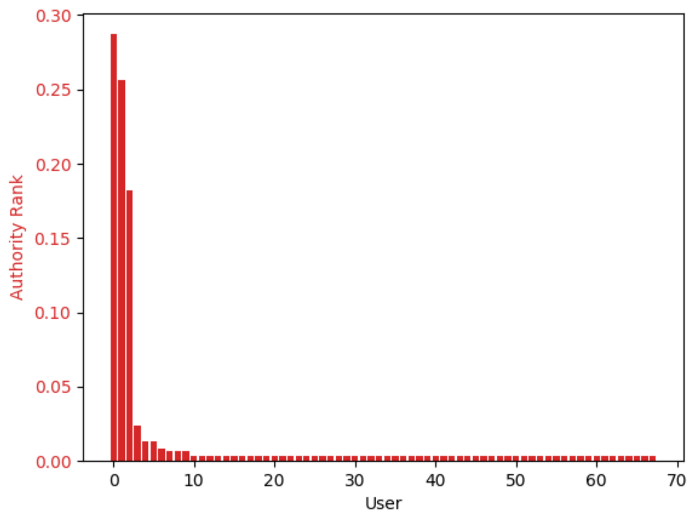

Figure FC5.1: Authority Ranks for Gratia sorted in decreasing order. Stable λ = 51

The first observations we make are of the computed Authority Ranks. Figures FC5.1 and FC5.2 shows the Authority Rank values for Gratia and CircuitVerse respectively while Figures FC5.3, FC5.4 and FC5.5 show the AR values for the simulated Barabasi–Albert, Random and Watts–Strogatz networks respectively. Somethings are immediately apparent. Amongst the real networks, CircuitVerse has a much more skewed

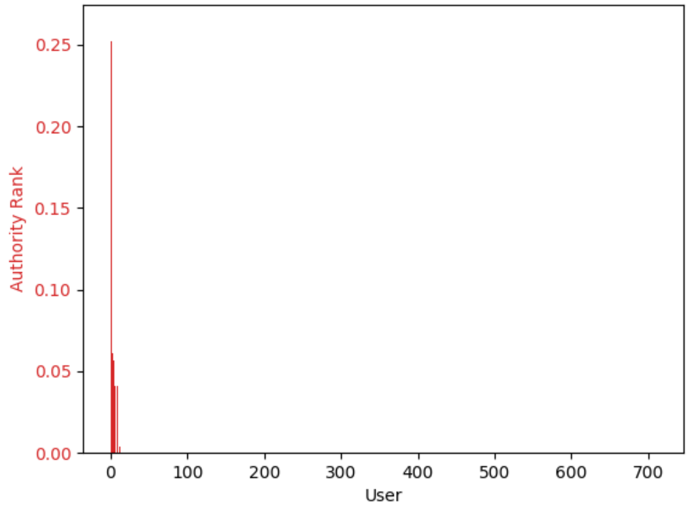

Figure FC5.2: Authority Ranks for CircuitVerse sorted in decreasing order. Stable λ = 35

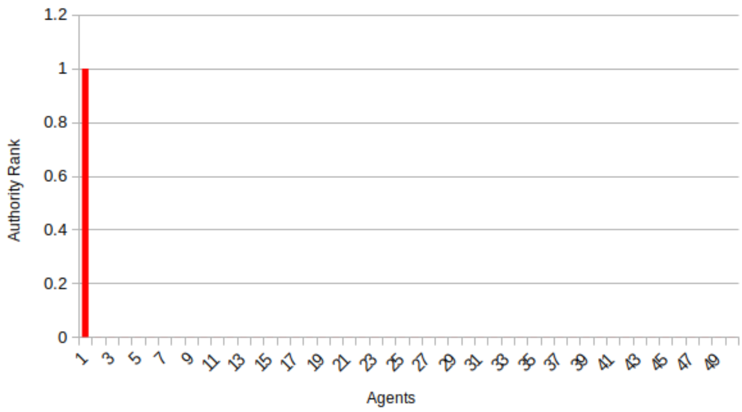

Figure FC5.3: Authority Ranks for a Barabasi–Albert network sorted in decreasing order. Stable λ = 12

distribution of AR values than Gratia.

Amongst the simulated networks, Barabasi–Albert has the perfectly skewed AR distribution, one agent has a AR of 1 and every other agent has 0. This means that it is the network's belief that the lone agent with AR = 1 can influence the entire network. Comparatively, Watts–Strogatz and Random networks are less skewed than Barabasi–Albert, with Watts–Strogatz (Figure FC5.5) being closest to flat.

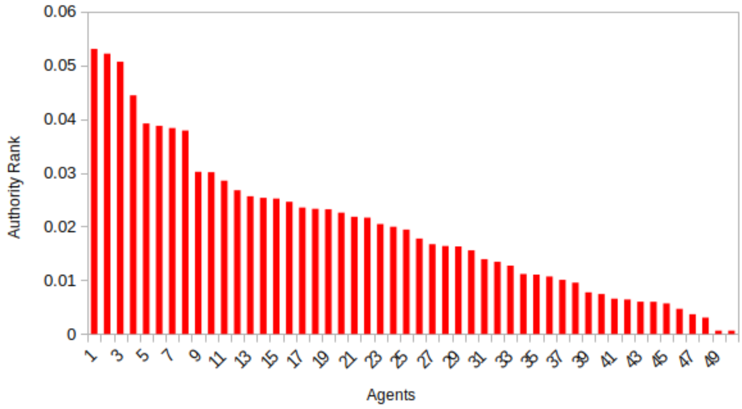

Figure FC5.4: Authority Ranks for a Random network sorted in decreasing order. Stable λ = 26

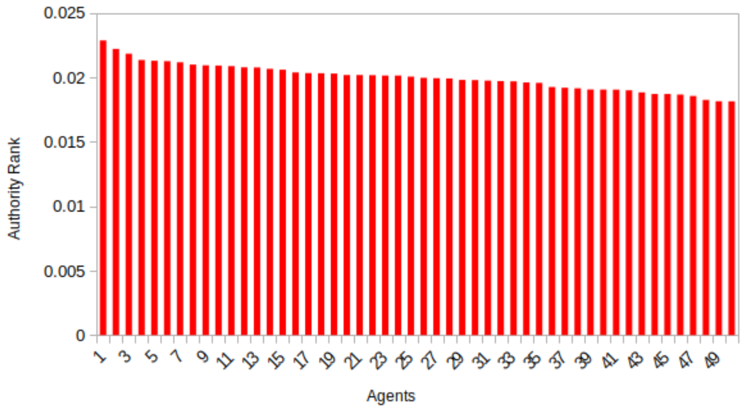

Figure FC5.5: Authority Ranks for Watts–Strogatz network sorted in decreasing order. Stable $\lambda = 4$

Further, AR distribution of Barabasi–Albert is the closest to the AR distributions of Gratia and CircuitVerse, thereby, suggesting that both are also a type of Barabasi–Albert network, albeit imperfect.

5.2 Citizen Ranks

Figures FC5.6 and FC5.7 shows the Citizen Rank values for Gratia and Circuit-Verse respectively while Figures FC5.8, FC5.9 and FC5.10 show the CR values for the

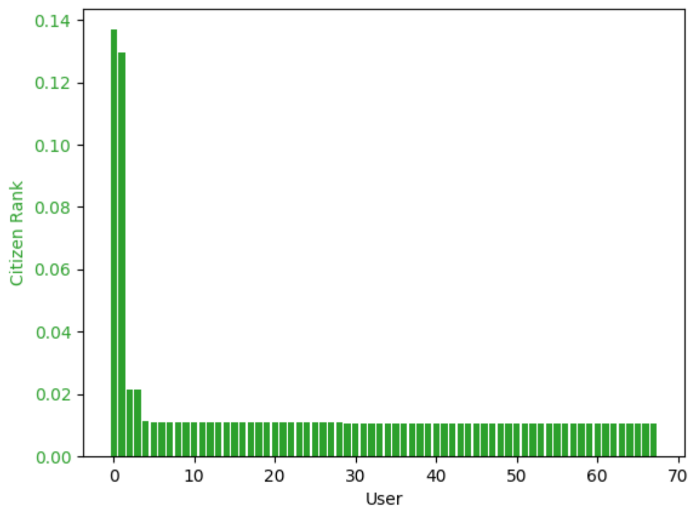

Figure FC5.6: Citizen Ranks for Gratia sorted in decreasing order. Stable $\beta = 46$

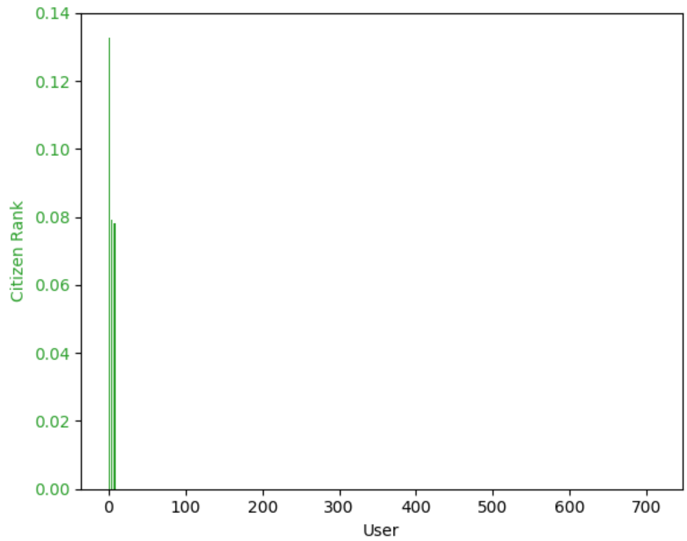

Figure FC5.7: Citizen Ranks for CircuitVerse sorted in decreasing order. Stable $\beta = 42$

simulated Barabasi–Albert, Random and Watts–Strogatz networks respectively.

Amongst the real networks, CircuitVerse has a much more skewed distribution of CR values than Gratia.

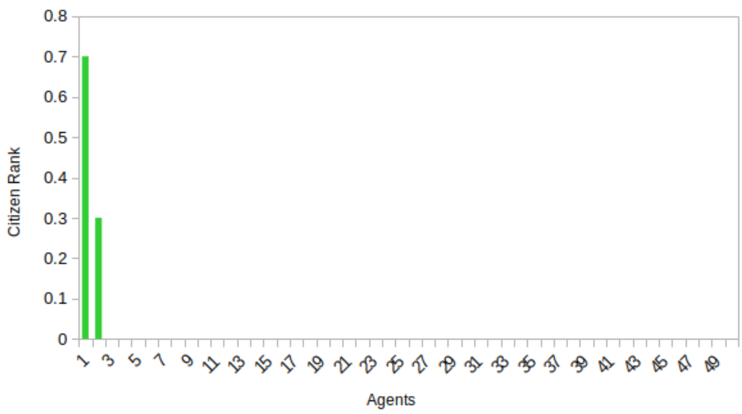

Figure FC5.8: Citizen Ranks for a Barabasi–Albert network sorted in decreasing order. Stable $\beta = 12$

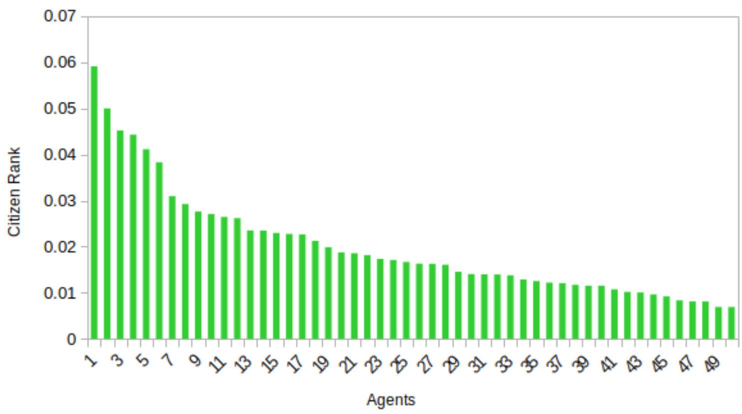

Figure FC5.9: Citizen Ranks for a Random network sorted in decreasing order. Stable $\beta = 11$

Amongst the simulated networks, Barabasi–Albert has the very skewed AR distribution, two agents together have a CR of 1 and every other agent has 0. This means that it is the network's belief that these agents contribute the most engagement to the entire network while others don't. Comparatively, CR distribution for Watts–Strogatz and Random networks look quite similar and are less skewed than Barabasi–Albert.

Again, the CR distribution of Barabasi–Albert is the closest to the CR distributions of Gratia and CircuitVerse, thereby, reinforcing the belief that both are also a type of Barabasi–Albert network, albeit imperfect.

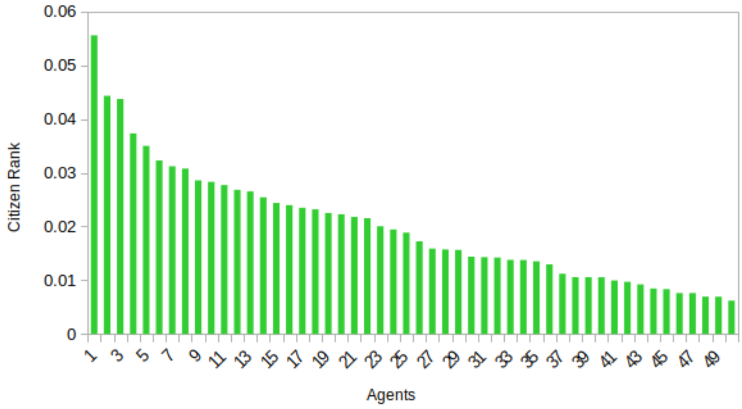

Figure FC5.10: Citizen Ranks for a Watts–Strogatz network sorted in decreasing order. Stable $\beta = 14$

5.3 Comparisons

5.3.1 Comparing across Networks

Hypothesis H1: Networks with lower average value of engagement will have a lower value of stable λ.

Table TC5.1: Comparing λ across networks

Network	**Avg Time per stroke**	λ
Gratia	6.06	51
CircuitVerse	4.49	35
Random	2.72	26
Barabasi–Albert	2.13	12
Watts–Strogatz	1.71	4

Table TC5.1 shows the correlation seen between average time per stroke and the stable point for λ. Gratia has a higher average time per stroke/interaction and also has the highest λ value. Comparatively, CircuitVerse has a lower average time and also a much lower stable λ value. This suggests that Gratia's users are more engaged with the other user's they interact with than CircuitVerse's users.

Even amongst the simulated networks, the Random network seems to have the higher average time and λ than Barabasi–Albert and Watts–Strogatz networks. This can be attributed to the fact that in the Random network, each agents connects with many other agents as compared to the other simulated networks.

The results from Table TC5.1, thus, confirm the Hypothesis H1.

5.3.2 Comparing with PageRank

Hypothesis H2: Authority Ranks will closely mimic PageRank for λ values which are much smaller than the stable λ .

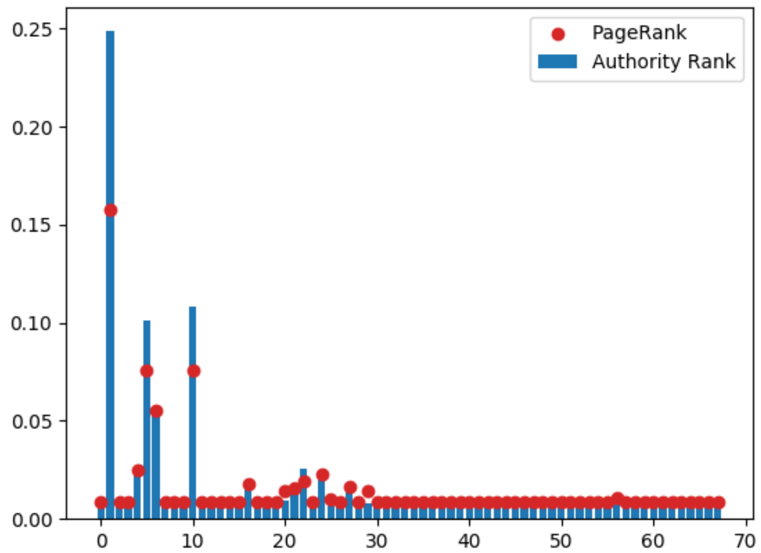

Figure FC5.11: PageRank and Authority Ranks for Gratia users with λ = 10

Figures FC5.11 and FC5.12 compare the PageRank for a user to their Authority Rank for the network of users of Gratia. When we supply a low value for $\lambda = 10$,i.e. for Figure FC5.11, the AR values closely mimic the PageRank values for all the users. The exact values might be different but the relative distribution is similar. On the other

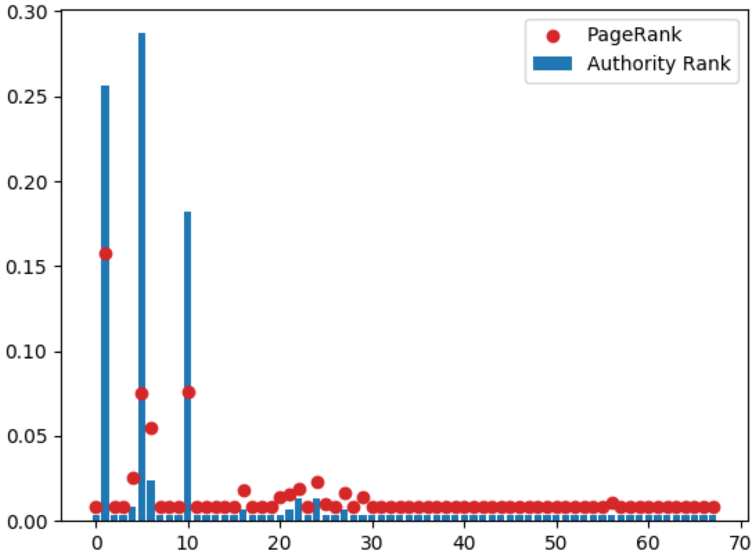

Figure FC5.12: PageRank and Authority Ranks for Gratia users with λ = 51

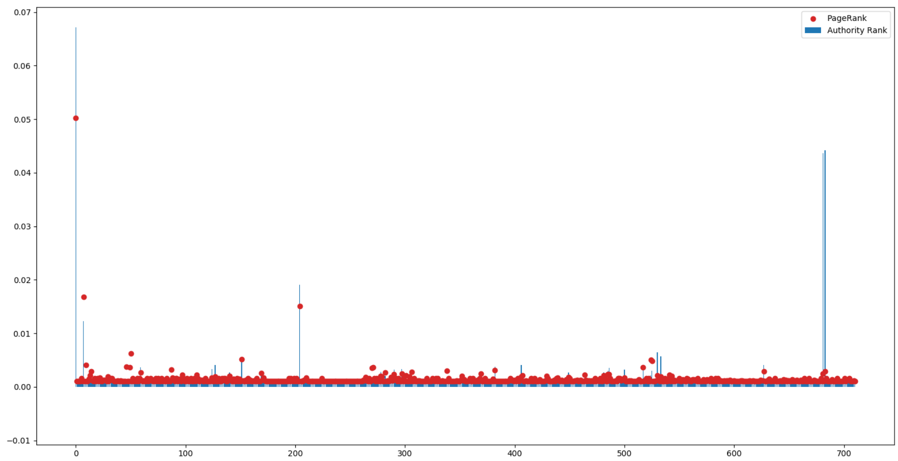

Figure FC5.13: PageRank and Authority Ranks for CircuitVerse users λ = 10

hand, for the observed stable value of $\lambda = 51$, i.e Figure FC5.12, the AR values follow a very different trend than the PageRank. Several users having the same PageRank, have different AR; while some users with low PageRank have a high AR and some other users have a high PageRank but low AR. This shows the stark difference be-

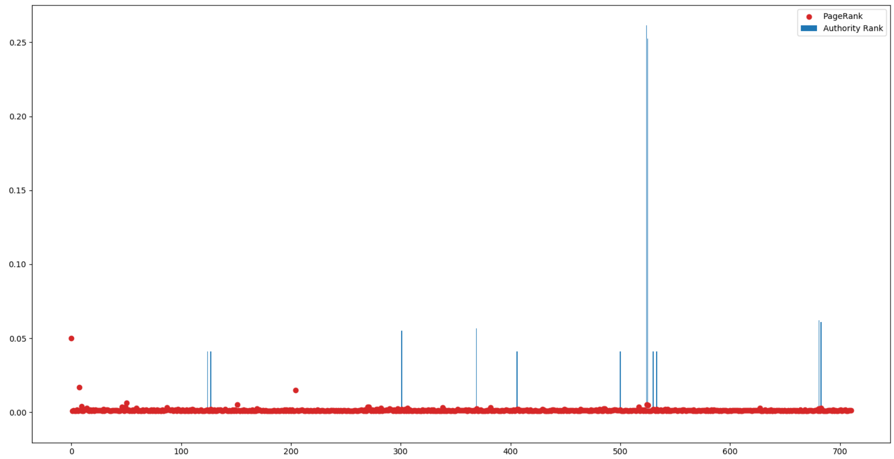

Figure FC5.14: PageRank and Authority Ranks for CircuitVerse users λ = 35

tween PageRank and AR. Users with high AR have engaged audiences who pay a lot of sustained attention, as compared to, users with high PageRank who have a larger audiences most of who give only cursory attention. PageRank fails to make the distinction between sustained and cursory attention, whereas AR does.

A similar trend can be seen for CircuitVerse.For a low value for $\lambda = 10$, i.e. Figure FC5.13, the AR values closely mimic the PageRank values for almost all the users, whereas, for the observed stable value of $\lambda = 35$, i.e Figure FC5.12, most users with high AR have very low PageRank and users with high PageRank have low AR.

These results, therefore, confirm Hypothesis H2.

CHAPTER 6

CONCLUSIONS

To the best of our knowledge, the concept of social capital and its importance in measuring social influence hasn't received enough research attention, particularly in the field of social network analysis. This work is an effort to address this gap and look at social capital and influence through the lens of engagement and belief revision and show that engagement is a much more reliable mechanism to measure influence than current mechanisms.

This work gives a model of engagement as a computational variable in form of two properties: scarce and renewable. Using this model and the concept of belief revision, two new network measures are proposed: Authority Rank and Citizen Rank.

Authority Rank is a influence measuring algorithm that takes into account the distinction between cursory attention and sustained attention and results from both simulated and real data show the same.

Citizen Rank is a new measure that give a score of the involvement of participants in the network. This too takes into account, the distinction between cursory and sustained attention.

The computational model proposed also gives us two observable variables: λ and β. The observed values for these can be used to make comparisons between different networks. λ gives a indication of how engaged the members of a network are, while β gives an indication of the value of the participation of its members.

Also, λ and β can be fed into the computational model and for different values different AR and CR distributions can be observed. By varying these parameters, different kinds of users in the network can be highlighted.

6.1 Applications

Authority Ranks can be used:

- To identify users in a social network who enjoy cult like engagement.
- As a more stable, grounded metric of measuring influence.

Citizen Ranks can be used:

- To identify users in a social network who are highly involved.
- To help content creators identify followers that engage and followers that don't.

6.2 Future Work

Some of the few ways this work can be extended is to use individual λ and β values for users. This would gives a more realistic insight as each user has a different way of engaging with others. These values would probably be estimated from similarities between users through their interest vectors.

One of the limitations of the this model is the computational effort involved which makes it difficult to implement for large networks. One of the ways to get past this limitations is to create topic networks from the original network and compute AR and CR values for that topic. Thus, the users would have a AR and CR score for each topic and their overall AR and CR would be vector or any other aggregate of these individual scores.

Bibliography

[1] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. Influence analysis in social networks: A survey. *J. Netw. Comput. Appl.*, 106:17–32, 2018.

[2] Hilde A M Voorveld, Guda van Noort, Daniël G Muntinga, and Fred Bronner. Engagement with Social Media and Social Media Advertising: The Differentiating Role of Platform Type. *J. Advert.*, 47(1):38–54, Jan 2018.

[3] S. P. Tussyadiah, Devi Roza Kausar, and Primidya K.M. Soesilo. The Effect of Engagement in Online Social Network on Susceptibility to Influence. *J. Hosp. Tour. Res.*, 42(2):201–223, 2018.

[4] Pierre Bourdieu. The forms of capital. In J Richardson, editor, *Sociol. Econ. Life, Third Ed.*, volume 8096, pages 78–92. Greenwood Press, New York, 2018.

[5] Michael W. Foley and Bob Edwards. Is It Time to Disinvest in Social Capital? *J. Public Policy*, 19(2):141–173, 1999.

[6] AleJandro Portes. Social Capital: Its Origins and Applications in Modern Sociology. *Annu. Rev. Sociol.*, 24(1):1–24, 1998.

[7] Robert D Putnam. *Making Democracy Work: Civic Traditions in Modern Italy*, volume 25. Princeton University Press, 1996.

[8] Robert D Putnam. Bowling alone: The collapse and revival of American community., 2000.

[9] Virginia Morrow. Conceptualising social capital in relation to the well-being of children and young people:. *Child Welf. Soc. policy*, 47(4):143–156, 2017.

[10] Charles Kadushin. Too Much Investment in Social Capital? *Soc. Networks*, 26(1):75–90, 2004.

[11] Ronald S. Burt. *Brokerage and closure*. Number 1. Oxford university press, 2014.

[12] Nancy Lin, Karen Cook, Ronald S. Burt, Barry Wellman, and Kenneth A. Frank. Network Capital in a Multilevel World: Getting Support from Personal Communities. In Nan Lin, Karen Schweers Cook, and Ronald S. Burt, editors, *Soc. Cap.*, pages 233–273. Aldine de Gruyter, New York, 2019.

[13] Martin Van Der Gaag and Tom A.B. Snijders. The Resource Generator: Social capital quantification with concrete items. *Soc. Networks*, 27(1):1–29, 2005.

[14] Ronald S. Burt. Structural Holes vs. Network Closure as Social Capital. In Nan Lin, Karen Schweers Cook, and Ronald S. Burt, editors, *Soc. Cap.*, pages 31–56. Aldine de Gruyter, New York, 2001.

[15] Nancy Lin, Karen Cook, Ronald S. Burt, Nan Lin, Yang-chih Fu, and Ray-May Hsung. The Position Generator: Measurement Techniques for Investigations of Social Capital. In Nan Lin, Karen Schweers Cook, and Ronald S. Burt, editors, *Soc. Cap.*, pages 57–81. Aldine de Gruyter, New York, 2019.

[16] N Lin. Building a network theory of social capital Cited by me. *Connections*, 22(1):28–51, 1999.

[17] Jack P. Gibbs and James S. Coleman. Foundations of Social Theory. *Soc. Forces*, 69(2):625, 2006.

[18] David Grusky and Mark S. Granovetter. The Strength of Weak Ties. In *Inequal. 21st Century*, pages 249–252. Elsevier, 2018.

[19] Amartya Sen. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philos. Public Aff.*, 6(4):317–344, 1977.

[20] Amartya K Sen. *Development as freedom*. Oxford University Press, 2017.

[21] Matteo Migheli. Capabilities and Functionings: The Role of Social Capital for Accessing New Capabilities. *Rev. Polit. Econ.*, 23(1):133–142, Jan 2011.

[22] Alexandre L. Bertin and Nicolas Sirven. *Social capital and the capability approach: A social economic theory*. Jan 2006.

[23] Isaac Mwangi and Shem Ouma. Social capital and access to credit in Kenya. *Am. J. Soc. Manag. Sci.*, 3(1):8–16, 2012.

[24] Veerle Vyncke, Bart De Clercq, Veerle Stevens, Caroline Costongs, Giorgio Barbareschi, Stefán Hrafn Jónsson, Sara Darias Curvo, Vladimir Kebza, Candace Currie, and Lea Maes. Does neighbourhood social capital aid in levelling the social gradient in the health and well-being of children and adolescents? A literature review. *BMC Public Health*, 13(1):65, Jan 2013.

[25] Kristian Bolin, Björn Lindgren, Martin Lindström, and Paul Nystedt. Investments in social capital - Implications of social interactions for the production of health. *Soc. Sci. Med.*, 56(12):2379–2390, Jun 2003.

[26] Tom Healy and Sylvain Côté. *The Well-Being of Nations: The Role of Human and Social Capital. Education and Skills.*, volume 47. ERIC, 2001.

[27] T. Semih Akcomak and Bas ter Weel. The impact of social capital on crime: Evidence from the Netherlands. *Reg. Sci. Urban Econ.*, 42(1-2):323–340, Jan 2012.

[28] John Humphrey and Hubert Schmitz. Trust and inter-firm relations in developing and transition economies. *J. Dev. Stud.*, 34(4):32–61, Apr 1998.

[29] Eric Berne. *Transactional analysis in psychotherapy: A systematic individual and social psychiatry.* Grove Press, New York, US, 2007.

[30] C. Christman. *Games People Play: The Psychology of Human Relationships*, volume 5. Grove Press, New York, 2013.

[31] Xin Yang and Ju Fan. Influential User Subscription on Time-Decaying Social Streams. *CoRR*, abs/1802.0, Feb 2018.