Sustainability of CAPS Social Network: a Network Analysis Approach using Agent-Based Simulation

Thesis Master Information Studies – Business Information Systems

Peter Gerbrands (11079800)
peter.gerbrands@student.uva.nl
University of Amsterdam, Faculty of Science

Final version
June 30, 2016

Supervisor	Second Examiner
Dr. Martin Ruivenkamp	Dr. E.J. de Vries

Abstract

This paper focuses on analyzing the structure of several egocentric networks of collective awareness platforms for sustainable innovation (CAPS). It answers the question whether the network structure is determinative for the sustainability of the created awareness. Based on a thorough literature review a model is developed explaining and operationalizing the concept of sustainability of a social network in terms of importance, effectiveness and robustness. By developing an agent-based model, the expected outcomes after the dissolution of the CAPS are predicted and compared with the results of a network with the same participants but with different ties. Twitter data from different CAPS is collected and used to feed the simulation. The results show that the structure of the network is of key importance for its sustainability. With this knowledge and the ability to simulate the results after network changes have taken place, CAPS can assess the sustainability of their legacy and actively steer towards a longer lasting potential for social innovation. The retrieved knowledge urges organizations like the European Commission to adopt a more blended approach focusing not only on solving societal issues but on building a community to sustain the initiated development.

Table of Contents

1 Introduction

The digital agenda for Europe, a European initiative under the Horizons 2020 program, shows that in the period from 2012 to 2017 in total € 79 million is allocated to Collective Awareness Platforms for Sustainability and Social Innovation (CAPS)[1]. According to (Arniani et al., 2014, p.9), the European Commission defines CAPS as:

> *"The Collective Awareness Platforms for Sustainability and Social Innovation (CAPS) are ICT systems leveraging the emerging "network effect" by combining open online social media, distributed knowledge creation and data from real environments ("Internet of Things") in order to create awareness of problems and possible solutions requesting collective efforts, enabling new forms of social innovation."*

The referenced "network effect" is a phenomenon in which the value of the product is influenced by the number of users of that product (Shapiro & Varian, 1999, p.19). So the more active participants a CAPS has, the higher the value of the system for the users (including the CAPS itself). Besides value for the users, as the definition of a CAPS includes, the goal is to create social innovations, which focus on a collective instead of an individual value. Creating this collective value, i.e. a collective good, is a complex process of group dynamics due to differences between individual interests and the interests of the group (Olson, 1965). An important factor that enables collective action in this complex process is heterogeneity (Oliver, Marwell, & Teixeira, 1985), since it brings together a high incentivized people with low incentivized people, who would otherwise not be activated at all. This implies, however, that a lack of heterogeneity intensifies the gap between participants and the rest of the society (Arniani et al., 2014, p.15), resulting in a more biased view of the problem at hand. Therefore generating the collective value of the CAPS depends on the diversity of the participants. Besides the heterogeneity of the network, the interdependence of participants, i.e. how participants influence each other; are crucial for an effective CAPS, since this "interdependence can yield a cascade of activism and result in a successful social movement" (Kim & Bearman, 1997).

To create collective value, most CAPS develop specific tools that utilize the knowledge of the crowd and serve as a central repository of information for the crowd, which is made available online. But what happens with these tools when funding stops? Who will continue to exploit, maintain and support these systems? This paper argues that a CAPS provides more value than just the tools they develop. By using diverse social media, e.g. Twitter, they have created awareness of the problems they deal with and by doing so created virtual communities around different fields of interest. Virtual communities transform society by integrating products and services, blending national identities, by integrating or fragmenting diverse communities, or by creating new personal relationships (Romm, Pliskin, & Clarke, 1997). So these virtual communities, which are represented as a social network, affect society

[1] See European Commission: https://ec.europa.eu/digital-agenda/en/collectiveawareness

and their heterogeneity forms the base for a realm of new opportunities and initiatives to further develop society, as long as they continue to communicate with and influence each other actively.

Analyzing the participants of a network, the extent to which they influence each other, the level of heterogeneity of the network, and how information flows between participants are just a few of the possible methods (social) network analysis provide. By analyzing network structures, how they affect the behavior of a society or community can be explained (Jackson, 2008, p. 3), which is what this research focuses on. By assessing the behavior of a CAPS community using the existing structure of its social network (the ego-network), the sustainability of the community can be simulated after removing the CAPS itself from the network. Therefore, the central question this research answers is:

> ***To what extent does the structure of an egocentric social network affect the sustainability of that network after dissolving the CAPS?***

To answer this question, this research defines the sustainability of a network in terms of *importance*, *effectiveness* and *robustness*. The importance of the network is defined as the ability of the network to affect reality. Participant will leave the network when it does not keep its members engaged, thus does not provide individual value, resulting in dissolving the network and its impact. The individual contributions cumulate to a collective value and when the network does not provide a collective value for society, there is no need to sustain it (in the scope of this study). While the perceived value is hard to measure, the outcome of this value is more easily assessed by measuring the activity. The effectiveness of a network is defined as how the structure of the network is able to affect its participants to increase their activity, ensuring a continuous growth of activity, developing the potential to form new initiatives. Finally, the robustness of the network focuses on assessing to what extent the structure of the network can be distorted while keeping its value and effectiveness.

Events of collective action are frequently analyzed and described using Twitter data, e.g. climate change (Segerberg & Bennett, 2011) or political turmoil (Christensen, 2011; Morozov, 2009). Arguing that social movements frequently adopt new ICTs and non-institutional channels, a more general framework for analyzing the Twitter content is provided (Bajpai, 2011). These studies explain the importance of Twitter for initiating collective action, the goal of a CAPS, recognizing its speed or volatility and non-institutional character. Therefore this study analyzes Twitter data too, but in contrast to the studies mentioned, not focusing on the content but on the egocentric networks that emerge via Twitter activity around the CAPS's.

First extant literature is reviewed to define and operationalize the conceptual model. After delineating the model, an agent-based model is developed that simulates the sustainability of the network. Using empirical data collected by studying existing CAPS initiatives and inserting this in a simulation, the sustainability of the egocentric network of each CAPS can be assessed.

The importance of this study lies in finding a significant relationship between the mentioned characteristics of the network and the resulting sustainability as well as the extent to which this relationship affects sustainability. If these are found, it can be presumed that CAPS can actively influence the structure of their network in order to increase the life-span of the network and thus ensure its legacy. Since open online social media are integral part of the definition of a CAPS, extending the sustainability of the network increases the probability of success and therefore its effect on society.

After exploring extant literature on the analysis of Twitter (and other social media networks) in section 2, the following chapters discuss the current body of literature and justify the conceptual framework. It identifies the concepts and measures available for assessing sustainability, including the associated limitations from the perspective of the sustainability of CAPS. First the importance of online social networks is discussed in section 3, demonstrating that these virtual networks have a significant effect on the “real world” which gives them their value for both the participants as for society. Section 4 focuses on how the participants affect each other, arguing that the structure of the network is of key importance and is strongly related to communities and similarities between participants. The last factor to determine a network’s sustainability is robustness, see section 5, arguing that even when a network is important and effective, when its structure cannot deal with perturbations, the network cannot sustain. The literature review is concluded by connecting the terminology in section 6.

This study is heavily based on concepts from the field of social network analysis. By employing agent-based modeling and inserting retrieved network data, it creates simulations that predict the sustainability given the expected models of influence, which are explained in section 7. Section 8 describes the data collection process and how it is processed. Next, in section 9, the developed simulation is explained, after which the results are presented in section 10. Finally section 11 presents the conclusions, the theoretical and practical relevance and the limitations of this work providing direction for future work on the subject.

2 Predicting Twitter: Content, Users and Trends

This paper focuses on predicting how the activity of a user influences the Twitter network, triggering other users to become active and explains the sustainability of the network over time. Extant literature classifies three types of predictions, in general focusing on the content, the user or the interaction type.

In content based models predictions on the popularity of hashtags are created by extrapolating time-series information using wave patterns (Doong, 2016) or by perceiving the popularity as a classification problem instead of finding an exact value (Ma, Sun, & Cong, 2013). While these models may provide insight into what motivates people to contribute, they ignore the effect of influencing others using the friendship structure.

Studies that focus on Twitter users typically include this structure and explain leadership of the user outside of the Twitter network (Xu, Sang, Blasiola, & Park, 2014), keeping users engaged using collaborative filters to proactively provide information relevant to similar types of users (Diaz-aviles et al., 2014). Another method for assessing the influence of users on Twitter is to compare them with existing rankings, after which some observations on Twitter networks are made (Nguyen & Zheng, 2014). Although these user-based models incorporate the friendship structure, how this structure affects influence is still undetermined.

A third type of analysis focuses on interaction types, answering questions on what type of interaction is feasible for what type of motivation (Alhabash & McAlister, 2014) or how a specific interaction (re-tweets) follows a power-law distribution which is used to predict further retweets (Lu, Zhang, Cao, Hu, & Guo, 2014). Only one article is retrieved that combines the three foci in explaining hashtag popularity (Zhang, Wang, & Li, 2013). Although much attention is given to Twitter and other social media, little is known about how the friendships within the network affect the value of the network, especially for egocentric networks. This is what this study aims to unravel.

3 Importance: Real-world Value of Social Media

While one could question the importance of social media in the “real world”, this section elaborates on the ability of social media to act in its environment. The next section provides an overview of literature arguing the importance or agency of social media. Because of this agency, the network created and sustained by the social media becomes valuable for both the participants as for society.

3.1 Real-world Agency

Mass media and the natural tendency of actors to compromise, influence the public opinion (Boudin, Salvarani, Boudin, & Salvarani, 2015). Although different social media communications affect public engagement and organization-public relationships (Men & Tsai, 2015) it still requires a better understanding from a user-construction perspective identifying different ways of perceiving value (Smith & Gallicano, 2015). Organizations use social media for CRM activities using diverse channels including content community e.g. YouTube and Flickr, crowdsourcing, microblogging i.e. Twitter, and social networking e.g. Facebook (Go & You, 2016). Other organizations design business models that use the knowledge of the crowd (Nik-Bakht & El-Diraby, 2015). It is undeniable that social media plays an important role in current communications. Considering this importance of social media in our current communication structure, it must also influence the public opinion (G. Wang, Liu, Li, Tang, & Wang, 2015), thus giving agency to social media platforms, transferring the effects from the virtual world to the physical.

Now that these platforms affect “real life”, these platforms are used for creating a sense of "belonging" and "being popular" (Chang, 2015) or in other ways contributing to individual personality traits

(Hollenbaugh & Ferris, 2014), thus affecting life satisfaction (Oh, Ozkaya, & LaRose, 2014). Although these factors indicate the personal value of social networks, the sustainability of networks is primarily assessed from an economic approach (Shapiro & Varian, 1999), while the non-economic approach, i.e. the mentioned personal values, is frequently disregarded in the calculation of the overall value proposition. An Ising model (adopted form physics explaining magnetism) and assessing historical events, potentially explains whether a new idea is supported (progressive +1) or not (conservative -1). Using this, a tool for predicting opinion shifts is developed that assesses peer-to-peer and mass communication patterns (Kindler, Solomon, & Stauffer, 2013). While this demonstrates the potential power that social network platforms have in reaching consensus (Jalili, 2012), the model is highly restricted in practical use due to its theoretical configuration parameters and the restriction on the fixed amount of ties that each „spin" i.e. person, has. The ability of like-minded actants to form a community is greatly expanded by recommender systems using social platforms (Baraglia, Dazzi, Mordacchini, & Ricci, 2013), giving agency to the inclusion of online social networks in CAPS. Although several user-related factors have been identified (confirmation, satisfaction, perceived usefulness, enjoyment and habit) to persist continuous use of social media (Mouakket, 2015), how these are affected among users and how they influence the collective value of CAPS is still underexposed.

3.2 Individual and Collective Value

If a user does not perceive any value in using a social network, he or she will not use the network, so what determines the value of the network from a user perspective? When the network diffuses useful information (Aizstrauta, Ginters, & Eroles, 2015) or a person is influenced by his social environment (Li, 2013) the acceptance and use of the network grows. Besides these, many other factors contribute to the value proposition, as it is perceived by potential users, though, research on the technology acceptance model (TAM) focuses primarily on individual motivations for using a system and assesses the influence between the different components of TAM (Davis, 1993; Fan & Suh, 2014). While the TAM is extended by explaining how autonomy, relatedness and competence influence acceptance (Lee, Lee, & Hwang, 2015), how life satisfaction stimulates continuous use of social networks (Oh et al., 2014), and how social influence and cognitive instrumental processes affect acceptance (Venkatesh & Davis, 2000), the network effect within the acceptance process is largely underexposed.

An important concept for assessing the network effect on the TAM is the notion of critical mass. While the role of critical mass in the TAM is explained (Rauniar, Rawski, Yang, & Johnson, 2014), it is identified as the "weight" a user has when using the social network and the related hypothesis is confirmed that the critical mass of the user affects how this user perceived the value of the network. Secondly, the critical mass is given as given fact bout the user and is not incorporated as a dynamic recurring factor in the model itself. The role of critical mass with respect to collective action is a popular research theme in older literature (Marwell, Oliver, & Prahl, 1988; Oliver et al., 1985; Olson,

1965) but it has not included social networks as mass media communication due to its inexistence at the time the referenced literature was written.

Other than using the TAM for assessing the (perceived) value of social networks, it is also assessed in terms of influence, frequency, relevance, uniqueness, distance and community size, but the same formula incorporates "value" as a variable in the numerator (Geddes, 2011). This means that by applying some mathematics the value can be identified as a function of the perceived value resulting in a catch-22[2] problem. Considering that the perceived value provides the incentive for people to contribute to the network, the stronger the incentive for people to contribute (the higher the individual gain), the more people will contribute (the larger the cluster) and the quicker the goal (i.e. the delivery of the public good) is reached. The incentive is defined as the level of obtaining the collective good, which is affected by the price a user pays for participation, thus it follows rules of price elasticity of demand (Olson, 1965, p. 25), thus it affects the group size.

While a larger group size seems profitable, it is questionable whether the required group size is manageable for a CAPS in a social network, since biologically, the brain is limited to managing a certain amount of active relations (Dunbar, 1992) which is later identified for humans to be around 150 (Purves, 2008), known as Dunbar's number. This number is confirmed in a study on Twitter based conversations, where the ability of a user to reply to conversations decreases after the amount of friends surpasses their natural capacity, being somewhere between 100 and 200 friends after which the user starts to selectively weigh the importance of their friends (Gonçalves, Perra, & Vespignani, 2011). Adding this weight means that as the number of friends increase, the increase in participation in conversations decreases. This is known as a linear threshold model in which monotonicity (adding elements always results in either increasing or decreasing the value function) and sub-modularity (each time elements are added the new value function results in a reduced effect) characterize the model and is argued to be a good approximation for value optimization functions within a non-competitive environment (Borodin, Filmus, & Oren, 2010).

After reviewing the literature on determining the value of a network reveals the complexity of actually determining this value. Several ambiguous concepts of value are discussed, which do not clearly support identifying how to determine the actual network value. While all referenced works provide some utility function in which costs or a threshold value is subtracted from the generated value, this study will not attempt to do the same, but looks at the outcome of the utility function instead. When the individual value of using the Twitter network, is higher than the associated cost, the system will be used actively thus tweets are generated.

[2] According to the Oxford online dictionary: A dilemma or difficult circumstance from which there is no escape because of mutually conflicting or dependent conditions.

Therefore let the individual (perceived) value of the network for participant i be denoted by V_i, then it is calculated by comparing the participant's average daily activity within the network $\overline{a_i^I}$ is compared to the participant's average daily activity on Twitter in general $\overline{a_i^E}$ according to the formula

$$V_i = \frac{\overline{a_i^I} - \overline{a_i^E}}{\overline{a_i^I} + \overline{a_i^E}}$$

where V_i ranges between [-1, 1] in which a negative number signifies that the participant is less active in the network than they normally are on Twitter (thus the network is considered less important). The collective value of the network is approximated by the median of the individual values, denoted by $\tilde{V}$ (instead of the mean because of its sensitivity to outliers).

3.3 Friends and Followers

A simple way to determine the perceived importance of the network by a member is by counting how many followers they have, i.e. the out-degree. Having a large amount of followers automatically makes the network more valuable for the member, because with the same effort the member can reach more people. Unfortunately, this metric provides a distorted view since members with many connections are more likely to have many followers. While this gives them greater power to influence the network, their intention might be different. This study assumes that when people want to influence the network, so their perceived individual value is high, they will have more followers compared to their total amount of connections than those who do not, hence the formula:

$$F_i = \frac{d_i^O}{d_i^O + d_i^I}$$

where F_i is the followers-ratio, the amount of followers relative to the total amount of ties of node i, d_i^O is it's out-degree (the amount of followers) and d_i^I it's in-degree (the amount of friends). By comparing the followers-ratio of the overall Twitter connections F_i^{EXT} with the followers within the network, F_i^{INT}, one can determine whether the member is more active in influencing the network than in general, thus is assumed perceive the value of the network higher than of other networks.

$$\delta F_i = F_i^{INT} - F_i^{EXT}$$

While this measure is used to estimate the perceived individual value of the network for the member, it is closely linked to the amount of influence a node has in the network, which is further discussed in the next section.

4 Effectiveness: Dynamics of Influence

While, it may be evident that in order for online social networks to become important, they need to affect the real world substantially, this influence must come from within the network first. Only if the participants are affected by the network, society can be affected. So to what extent can a network affect its members? Therefore, this study refers to effectiveness of a network as the amount in which the network structure allows each participant to affect its neighbors and the total effect of this influence. Of course the level of influence is not the same for each participant, some people trust each other more than others and some are influenced less by others. These effects are part of the emergence of the network structure, thus the structure should reflect the influence levels accordingly. In social network analysis the type of influence in which similar people attract each other is called homophily and is detectable by identifying clusters.

4.1 Effectiveness as measure of influence

As the previous section demonstrated, participation in social networks is not only based on rational decisions, subjective factors like the cognitive and social needs of users, play an important role in awarding value to the network, which shows that people are influenced by their online social environment. In this study, the extent to which two participants influence each other is referred to as the effectiveness of the tie. In a Twitter network the follower-friend connections are directed and can be created without consent, meaning that one participant can follow another without prior consent, unless the other user blocks this person actively. Creating a tie, gives the first user (follower) access to the tweets of the latter (friend), but not vice versa. Once the tie is created, the follower is exposed to the activity of the friend and thus can be influenced, but in what direction and to what extent?

It seems logical to assume that if friends influence each other in real life, virtual friends influence each other in virtual life. Yet, only one study is found where the connections between users are linked with their actions in online open social media, exploiting homophily to explain the correlation (Yu & Xie, 2014). Unfortunately, the algorithm that Yu et al. have developed identifies social connections based on behavior while this study aims on explaining behavior based on social connections. Therefore the importance of homophily for explaining the correlation between social ties and behavior (influence) is proven but the endogeneity problem continues to exist. Besides seeing homophily as an indicator for alignment between actions and ties, when the participants (of a CAPS) pursue a similar collective good, users actively try to influence each other to induce homophily (Kim & Bearman, 1997).

A common function that determines the amount of influence or imitation within a social network, using either a discrete or continuous time, is the DeGroot model (DeGroot, 1974). Stating a simple version of the discrete model; the normalized effect that vertices in a network have on each other is presented in a matrix. Multiplying the matrix with the initial belief vector leads to a new belief vector which is the input for the next timeframe, and so on. If the set of nodes is strongly connected and closed they can either converge or be aperiodic (Jackson, 2008, p. 233). When the function converges, so at some point in time multiplying the belief vector with the matrix results is the same belief vector, no changes occur in the level of influence, resulting in a stable situation in which the activity for each user in the network will grow or die and the total amount of activity in the network will reach its limit.

Although the DeGroot model is simple and intuitive it assumes a priori knowledge about the level of influence of the individual users in order to calculate the overall outcome or consensus. An alternative for the belief vector which is often used in social network analysis is to focus on the eigenvector centrality C_i^E of each vertex i (Jackson, 2008, p. 41). Although the real belief vector is not available and the eigenvector centrality cannot be perceived by users directly, the introduced concept of homophily is important in estimating the level of influence.

4.2 Trust and Caution with Homophily

As mentioned in the previous section, homophily is the key to the alignment of ties between people and their behavior. Homophily characterizes how networks form, thus its structure and how it influences behavior and is determined by geography, family ties, interests and focus and diverse cognitive processes (Mcpherson, Smith-lovin, & Cook, 2016). In this study, the homophily of a group is defined as the ratio of the amount of incoming connections from nodes within the group to nodes within the group and all the incoming connections from all the nodes within the group. Let h_k define the homophily of group k and N_k defines the set of nodes that are part of group k. Let the total in-degree[3] of a node i be d_i and the amount of incoming connections from within group k is denoted as s_i than the homophily of group k is (Jackson, 2008, p. 19):

$$h_k = \frac{\sum_{i \in N_k} s_i}{\sum_{i \in N_k} d_i}$$

where the range of $h_k = [0, 1]$ in which the closer the value of h_k is towards 1, the higher the amount of homophily, i.e. the more in-group connections the nodes have.

While homophily explains why influence takes place, it does not quantify the influence itself. Therefore, this study combines the DeGroot model with the measure of homophily and instead of weighing each tie equally based on the in-degree of the node at hand, the source of the edge plays an important role. When the tie comes from a vertex within the same group, it is weighed differently than if not, using the homophily indicator h_k mentioned before. Let the weighed impact of an incoming edge from node j to node i be denoted as w_{ij} than the value is determined by:

[3] The original function does not use directed edges thus there is only a degree and no difference was made between the in-degree and out-degree.

$$w_{ij} = \begin{cases} j \in N_k: & \dfrac{1}{d_i}(h_k) \\ j \notin N_k: & \dfrac{1}{d_i}(1 - h_k) \end{cases}$$

where d_i is the out-degree of node $i \in N_k$, the set of vertices in group k. While the neighbors of a node affect the node, there is always a sense of self preservation, a measure of self-importance, denoted as w_{ii}. This measure is not assessed but is set as a constant in the model, ranging from [0, 1] in which 1 signifies that the node is omnipotent and will not be affected by its neighbors at all.

This study evaluates two different types of influence, influence caused by the amount of activity of a vertex's neighbors (which is easily perceived by the Twitter user) and one based on influence at a perceived value level. The reason for incorporating the second measure is based on the assumption that someone who does not use Twitter frequently will be affected by the neighbors' activity but will not change its level of activity dramatically, but will increase its level of activity as appropriate for its overall Twitter activity. The new activity of vertex i, a'_i, influenced by the activity of a neighbor (unless omnipotent) by:

$$a_i' = w_{ii}a_i + (1 - w_{ii})w_{ij}(a_j - a_i)$$

where a_i is the current activity of i, a_j is the activity of vertex $j \in n(i)$ (a neighbor of i). Alternatively the node could be affected by the individual perception of the network value by its friends, resulting in a different analysis. The value of the node depends on the weight given to its neighbors (including homophily) and its tendency to trust itself, thus:

$$V'_i = w_{ii}V_i + (1 - w_{ii})w_{ij}V_j$$

where v'_i is the new perceived value of node i, v_i is the previously perceived value and the weights remain unchanged. After calculating the new value, the associated activity can be calculated by adjusting the original perceived value function V_i to:

$$a_i^I = \frac{(-V_i - 1)a_i^E}{(V_i - 1)}$$

These formulae depict a discrete-time model and thus will change over time. Although it is interesting to find equilibrium points (the stable state after which the network values do not change), calculating these potential equilibrium points is impossible since they depend not only on the change in activity of all the neighbors but also on the sequence of calculating the activities. Therefore this formula is only useful in the simulation and running the simulation multiple times results in slightly different outcomes. As mentioned in the previous section, the effectiveness of the network describes the extent to which the vertices affect each other. Having created an algorithm on how to determine the amount of influence users within the same group have versus those outside of the group, the next step is to determine how to identify the groups themselves.

4.3 Community Detection by Clustering

There are many algorithms that detect clusters or communities in a social network, usually developed for specific needs. In general there are four different ways of determining cliques based on the mutuality of ties, the closeness or reachability of subgroup members, the frequency of ties among nodes and eventually based on the relative frequency of ties between members and non-members (Wasserman & Faust, 1994, p. 251). Since Twitter adopts a directed network approach, thus one can follow (and be influences by) another without consent, the reciprocation of ties i.e. the mutuality of ties approach is not a feasible candidate. Also the distance between two members is irrelevant in a Twitter network since one can only be affected by another if they have a direct tie (so the distance must be 1). The third clique approach which is based on the frequency of ties between nodes can provide a distorted view since this mechanism leaves out nodes that have a low amount of ties in general while they are still part of the network and can still influence others (especially when they are the small but different ones who influence the critical mass as previously discussed). Therefore the best approach for identifying cliques is using fourth approach, determining the amount of ties between members and non-members.

An alternative to looking at the individual nodes to determine clustering is to look at how properties of the cluster change when adding nodes to the cluster. These methods often require a predefined amount of clusters to identify, which is unsuitable for this study. If the number of ties a cluster has with other clusters reduces when adding (or moving) a specific node (from one) to the (other) cluster, the node is changed. Unfortunately these methods are proven to perform weakly in real-life community structures (Newman, 2004).

To avoid having to pre-specify the amount of clusters, the strong p-cliques approach can be used, which is provided in Pajek (Mrvar & Batagelj, 2016). This method is feasible since it is designed for working with directed networks and a linkage proportion parameter can be set. This parameter identifies the amount of connections that needs to be part of the same clique for the node to be part of that clique. For example, when it is set to 50% the assumption is that when a user has more than half of its connections in the same clique, the users is part of that specific clique. When in real life there are many cliques, the chance of having 50% of the connections in the same clique is too high and will most likely never occur, thus the network becomes one big clique (which is the only configuration that meets the requirement). Although the number of cliques does not have to be pre-specified, this clustering method still heavily relies on a-priori knowledge of the expected clusters.

Two other approaches that have gained popularity due to their inclusion in the popular social network analysis software Pajek (Mrvar & Batagelj, 2016), have not been tested for representing reality

(Newman, 2004), which are the Louvain method and the VOS detection method. While the Louvain method is originally developed for undirected networks and creates a hierarchical community structure that is optimized for computational speed and network size (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), it has been adapted to handle directed networks as well. The authors confirmed that the accuracy of detection remains similar to the traditional methods, but the time it requires for larger networks is significantly less. The VOS clustering technique is designed specifically for bibliometric networks (Waltman, van Eck, & Noyons, 2010) which focuses primarily on citation and co-authorship networks and co-occurrence of terms (networks that relate articles to terms). Because this study aims at the friend-structure of a Twitter network and disregards the content of the tweets, the VOS technique does not match the purpose and the Louvain method is used.

5 Robustness: Dealing with Disturbances

As the previous chapter describes, the structure of the network determines its effectiveness because during network formation, people tend to stick together when they feel they share a common perspective or beliefs. Since the network formation process is a continuous process, what happens to the effectiveness of the network and thus its importance, when the structure changes? To what extent can a network continue to provide value and thus opportunity for social innovations when its structure is stressed? Besides changes in the structure of the network, having different types of actors, i.e. a high level of heterogeneity, is of crucial importance for the probability of action. Only when enough members are engaged, or the critical mass is reached, this action will sustain and becomes collective action.

5.1 Scale-Free Networks and Percolation Theory

A common characteristic of many natural networks is that their degree distribution (i.e. how many connections each node has) follows a scale-free power-law. Instead of assuming that connections are formed randomly, i.e. the foundation for random networks, the scale-free network assumes that the connections between nodes are not randomly created but chosen specifically, known as preferential attachment (Barabasi & Albert, 1999; Barabási & Bonabeau, 2003). By simulating the removal of nodes in a network one comes to a critical point where the information within the network cannot reach all members of that network. This is a known application of percolation theory, showing that within a scale-free network, the level of robustness quickly reaches a state in which the network will not disintegrate assuming that the network has an infinite size (Cohen, Erez, Ben-Avraham, & Havlin, 2000).

While percolation theory provides an estimate for the critical level of nodes to be removed prior to the disintegration of the network, the method assumes random removal of the nodes. When nodes are removed selectively, i.e. the most important nodes are removed first, the network quickly deteriorates.

Nodes that play an important factor in connecting the network typically have a high betweenness centrality (Jackson, 2008, p. 39), thus the higher the centrality measure and the more nodes have these high values, the more fragile the network becomes.

An alternative approach that does not use random removal of nodes, uses entropy of the degree distribution instead to determine the networks stability (B. Wang, Tang, Guo, & Xiu, 2005). While in the developed simulation, the percolation theory is used, it is important to understand that heterogeneity is a key factor in the networks' robustness.

5.2 Heterogeneity as Access to Information

Having provided a function that predicts the users' activity, the impact on the collective value needs to be determined next. As mentioned earlier, the size of the group is of importance in collective action, but has an upper bound due to the characteristics identified as monotonicity and sub-modularity. The larger the group, the larger it's influence and the opposite applies to users with a large amount of friends. Although users with more friends than Dunbar's number (Purves, 2008), communicate more, they are likely to be less capable of participating in conversations. Therefore this study includes the amount of clustering (based on the amount of groups and their size) and the amount of friends as important factors for the value of the users' contribution to the network. While clustering can indicate the presence of heterogeneity, extant literature provides different definitions and explanations.

A network requires a relatively small amount of people that are willing to provide a large contribution to achieve a critical mass. These people usually diverge from the majority because they have an exceptionally large interest in the collective good or have access to a large amount of resources, i.e. are willing to make the costs (Oliver et al., 1985). Key to success for collective action is to have distributions of interest and resources containing a low average and large positive skew, and a positive correlation between the interest and resource availability must be present (p. 529-530). Although these conditions are tested successfully in many traditionally based settings, in a Twitter network, the cost function and the type of resources that provide value are very different. A better definition of heterogeneity in this paper is therefore defined as having access to (unique) information. To access information, the resource of importance is being connected to people who are not connected to the CAPS network, called external resources. For this reason, the heterogeneity of each user is based on the amount of fiends within and outside of the network. Using a normalized version of the "index of variation" (Agresti & Agresti, 1978, p. 206-208) the heterogeneity is calculated by:

$$I = \left(1 - \sum_{i=1}^{k} p_i^2\right) / \left(1 - \frac{1}{k}\right)$$

where $I = [0,1]$ is the normalized index of variation. Let k be the number of different categories and p be the proportion of observations in the ith category $(i = , \ldots, k)$. $I = 1$ means that the probability of a random link from a specific user is connected to the external resource is 50%, the highest possible heterogeneity level. While assessing the level of clustering within the egocentric network can identify a second level of heterogeneity, this internal clustering is already used for calculating the homophily level and thus gives too much importance to the internal clusters if used again. Therefore it is further excluded from the analysis.

5.3 Stability as Critical Mass Threshold

The importance of critical mass as mentioned in the beginning of the previous section is also visible in literature on the stability of social networks. From a societal perspective, the critical mass is assumed to be a penetration rate of 15% of the whole population (Geddes, 2011), which is used to argue for the importance of the introduced perceived user value function. While the paper provides a clear categorization of the different types of value (i.e. information, emotional, temporal and financial capital) it fails to provide any clarity on the contents of these types of value and to what extent they impact the perceived value. Secondly, their conclusion states "*Success is, however, absolutely reliant on getting it right first time*" (Geddes, 2011, p. 127), thus their critical mass is merely a measure of adoption rate similar to the diffusion of innovation (Rogers, 1995). Critical mass generated by a self-reinforced system, as perceived in this paper, also plays an important role in the stability of a network (Centola, 2013). The authors argue that if the incentive is too high – in contrast to the price elasticity of demand as mentioned in the previous section – the amount of users will grow rapidly and the stability of the system will deteriorate. Centola argues that although a weaker incentive results in a lower fraction of the population to participate (i.e. the free-rider problem), the expected point of critical mass lies further way from the total participation, thus the system can handle perturbation (i.e. disturbances) within the threshold between the critical mass a full participation level. While clear mathematical definitions are provided, the scaling parameters that identify the behavior are assessed theoretically and have not been given a real-life meaning and can vary between cases. Although these parameters can be calculated after collecting enough empirical data, for the topic of this paper, this is considered reverse engineering, i.e. making the model fit the reality of a single instance.

The literature demonstrates the relationship between critical mass and the stability of the network. When enough people are actively involved in the CAPS network (critical mass), disturbances (e.g. people leaving the network or becoming inactive, changing relationships between the participants) will less likely influence the stability of the network. This is because under weak incentives, so the amount of activity of neighboring participants (i.e. Twitter friends) influence the follower less, the likelihood for the system to be stable is higher since changes have a relatively low impact. Knowing that a weakly reinforced system is more stable than a strongly reinforced system and the first usually has a substantial amount of free-riders, an indication of stability is the proportion of free-riders (inactive network members):

$$f = \frac{(n - a)}{n}$$

where f is the proportion of free riders, n is the amount of followers and a is the amount of active followers (so $n–a$ is the amount of free-riders). The assumption in this study is: the larger the amount of free-riders, the weaker the incentive and therefore the more stable the network.

6 Sustainability of a Social Network

Sustainability is an ambiguous concept coming from biology where sustainability is "*originally understood as centering attention on the environment as a biological system that is able to endure and remain diverse.*" (Arniani et al., 2014, p. 10). It has been transformed to a more social approach focusing on the improvement of economic, environmental or societal aspects while not affecting the others (e.g. Nik-Bakht & El-Diraby, 2015) or it is used in a purely economic argument (e.g. Anand & Sen, 2000). Focusing primarily on the sustainability of a social network, sustainability has been defined in terms of the sustained membership within a community, focusing on long term presence of members (Gruzd, Wellman, & Takhteyev, 2011). As substantiated in the previous chapters, this study combines the approaches by defining sustainability in terms of *robustness* (if the presence of members changes the network will not fall apart), *effectiveness* (a more biological approach in which the total amount of influence that the network exerts on each user sums up to a positive amount, so the ecosystem is not depleted from its resources), and *importance* (the economic approach in which each member perceives individual value and the whole network provides collective value) as shown in Figure 1.

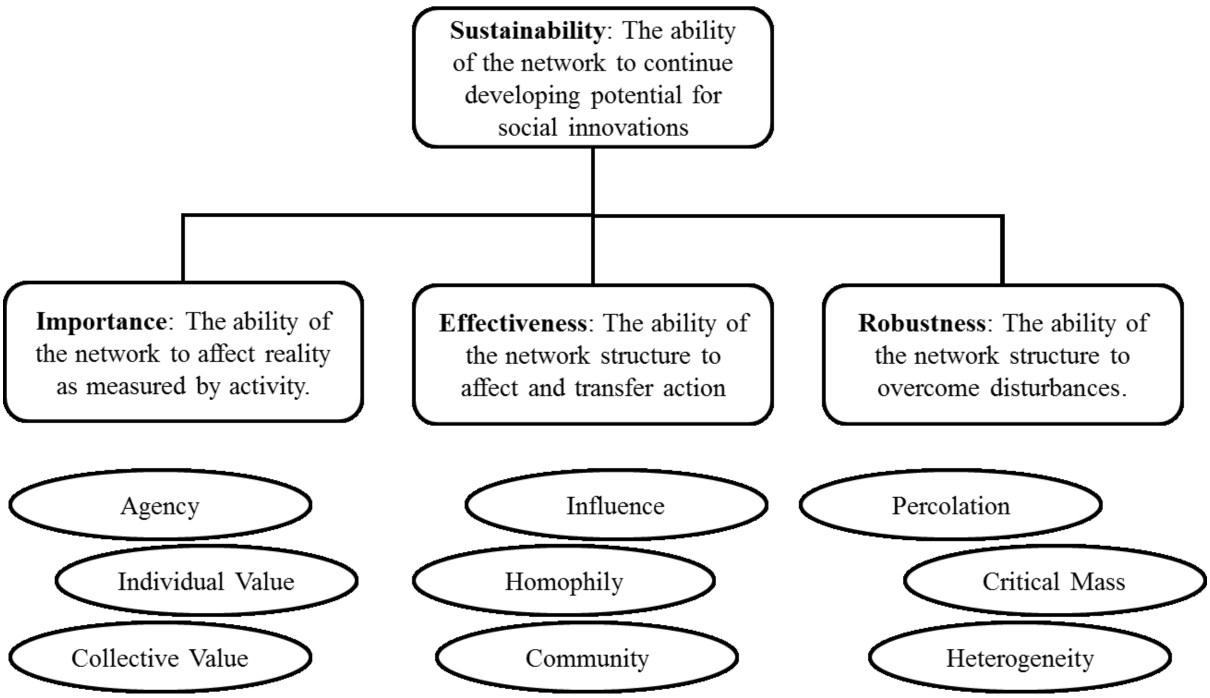


Figure 1: Conceptual model of Network Sustainability by author

Simply stated, a network is sustainable if it continues to provide value (the network shows a continuous flow of activity), the network affects the participants positively (the average of the weighted influence of the network is positive) and the network structure is robust enough to deal with random failures (removing nodes from the network does not directly affect the value and/or effectiveness of the network).

7 Method Selection and Execution

The research conducted consists of several sequential and concurrent steps. First, by means of a literature review, a conceptual framework is developed for assessing the sustainability of a social network. The retrieved knowledge is then converted into algorithms on how to assess the networks' sustainability. These algorithms are documented and implemented in an agent-based model that is used for simulation purposes. Finally actual data on the social networks of CAPS is retrieved, which then is inserted into the simulation to predict the future of the CAPS's network in terms of its sustainability.

Since the purpose of this study is to explain to what extent the structure of a social network determines the sustainability of the network, the structure is perceived as the independent variable. The dependent variable reflects the sustainability. Since the best indicator of the sustainability is the daily total amount of activity that the network generates over a longer amount of time, this amount is perceived as the dependent variable. All other variables used must remain constant. Since the outcomes of the agent-based simulation are dependent on the sequence of execution and a level of randomness, the results will differ after each run. Therefore the following hypotheses are defined:

> H_0 : *the average daily activity (or alternatively perceived value) of the real network is similar to the random network*

> H_a : *the average daily activity (or alternatively perceived value) of the real network is different from the random network*

By running the simulation 5 times for each network (if the variance is between the simulations is large, more simulations can be executed to generate a larger dataset), the randomness of the simulation is captured for both the real and random network. By comparing the outcomes using a standard t-test, the hypothesis is tested. Alternatively the same test is performed for the average perceived value.

7.1 Methodological Decisions

The primary research question focuses on answering how the structure of a social network affects the sustainability of a CAPS. Because the research starts with developing the conceptual framework and its operationalization, a multi-strategy design known as sequential transformative design is used

(Robson, 2011, p.165). By conducting an extensive literature review, on the existence of similar work and exploring of the phenomenon of sustainability in terms of importance, effectiveness and robustness of a social network, i.e. the qualitative analysis, a conceptual framework is developed. The conceptual framework is then operationalized in terms of mathematical formulae and variables, as explained in the appropriate chapters.

The temporal effect of the influence in a network makes the result from a mathematical viewpoint unpredictable. For example, if two participants are connected reciprocally, so they follow each other on Twitter, and their activity of perceived value differs to a large extent, which participant will adjust its activity first and how would that affect its neighbors? This study therefore approaches a CAPS network as a complex adaptive system (CAS) due to, its self-organization, the importance of heterogeneity (variability) and the indeterminacy and complexity of the interactions between participants (Harris, 2007, p.21-22). Because of these properties, a CAS often contradicts with the expectancies (created by a static model) and therefore the simulation using intelligent agents, an agent-based model, is a more feasible candidate (Sayama, 2015, p.19; Wilensky & Rand, 2015, p.5).

While it is impossible to internally validate the model using real data, since none of the CAPS stopped operating yet, an alternative approach is to compare the results of the simulation with real data against randomly generated data. By restructuring the network edges randomly and running the newly created random network (Jackson, 2008, p. 77) in the agent-based model, it is possible to identify the extent to which the network structure has affected the overall outcome. Using the simulation approach, allows for creating virtual experiments when real experiments are not possible, by altering the parameters and recalculating the network development (Wilensky & Rand, 2015, p. 335), enabling the grounding of advice on further developing the CAPS network.

7.2 Logical Considerations and Assumptions

The agent-based model is developed using NetLogo 5.3.1 (Wilensky, 1999) because of its strong and easily understandable programming language. But before the actual programming starts, several basic decisions are made.

The first complication is to determine the unit for the discrete-time model (NetLogo calls this a tick). In order to simulate the real activity as closely as possible the maximum average daily activity per user is used as time unit. For example, if the maximum user activity consists of 10 tweets per day, one tick represents 1/10th of a day. The time unit therefore is not similar for the different simulations.

Next, the process of influence needs to be specified. Two different processes are available, input-based or output-based. The input-based approach reviews per vertex its friends and updates the vertex based on all its incoming information. The second approach is output-based and updates all followers of an active node. The first approach would assume that the Twitter user reviews the activity of all its friends and then changes its behavior accordingly. The latter is chosen since it better reflects the use of Twitter, a user sees one or more tweets and decides to respond (or not), thus every tweet affects the user only slightly.

In section 4.2, two potential influence models are defined. A user can adjust its daily activity based on the amount of activity received from friends. This type of influence is easily detectable by the user but would assume that there is no personal limit to the amount of tweeting. Another possibility is for the user to create an image of the perceived value by its friends and adjust its own perception accordingly. As a result the user will change its behavior, i.e. activity, but always relative to its general Twitter behavior.

The agent-based model is based on a static network structure instead of a dynamic one. Although it is possible to generate new ties, add new users or remove them randomly or via preferential attachment (Jackson, 2008, p. 130), this will change the structure of the network making it complicated to isolate the effect of the structure itself on the sustainability of the network. As a result of the static network, the measures of robustness, heterogeneity and free-riders, will not change during the entire simulation. Therefore a linear regression approach is sufficient to indicate the effect of both factors on the sustainability of the network, despite the statement in section 7.1 that a static model does not predict accurately the evens in a complex adaptive system.

8 Data Collection and Operationalization

Based on two calls, 34 projects have been funded by the European Commission; the first call funded 12 projects[4] and the second call funded 22 projects[5]. Because the projects funded by the European Commission's second call are very recent, thus immature, they have been excluded from this study. For all projects funded by the first call, the details of the networks they host on Twitter are retrieved using NodeXL (Social Media Research Foundation, 2014) and the R programming environment (R Core Team, 2016).

In order to collect the appropriate data, the followers of the CAPS of interest are retrieved from Twitter using an script that loads the TwitteR (Gentry, 2015) package for R (R Core Team, 2016), resulting in a first zone ego-network (Knoke & Yang, 2008, p. 13) also known as "1.0". For each of the followers, the identifiers of their followers are retrieved (more data is not required and results in a huge increase in processing time) resulting in a second zone ("2.0') ego-network using another R-script that uses the RTwitterAPI (Vogler, 2014) and rjson (Couture-Beil, 2014) packages.

[4] Overview projects call 1: https://ec.europa.eu/digital-single-market/caps-projects-FP7
[5] Overview projects call 2: https://ec.europa.eu/digital-single-market/en/news/22-new-caps-projects-horizon-2020

Next, the detailed information and activity within the network of each of the retrieved followers is retrieved using NodeXL (Social Media Research Foundation, 2014) by importing the activity network using the retrieved follower screennames. Some basic characteristics are calculated, like counting the number of tweets each user has contributed to the network and calculating the average amount of daily activity for each of the users in general (all Twitter activity) and within the network (only activity between selected members). Since NodeXL does not retrieve the Twitter ID's of the nodes, they need to be matched with the retrieved data from R.

Next the retrieved data from NodeXL is loaded into Gephi (Bastian, Heymann, & Jacomy, 2009) and joined with the data retrieved from R. Because this study focuses on the post-mortem scenario of the CAPS, the ego, i.e. the CAPS itself, is removed from the network prior to performing the calculations, resulting in the final ego-network, named "1.5". An example of each ego-network level is shown in Figure 2.

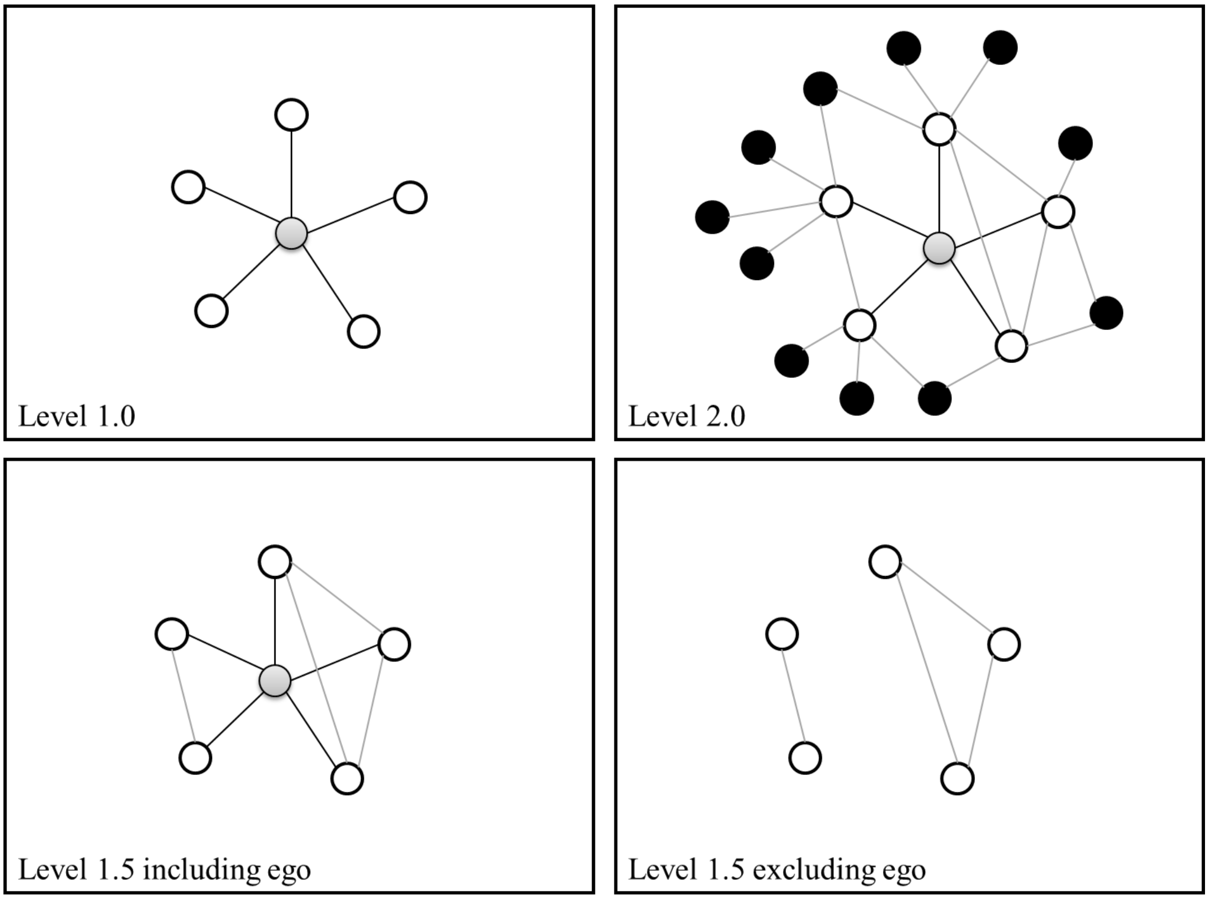


Figure 2: Overview Different Ego-Network Levels

Once the network is created in Gephi, some basic network statistics are calculated, like the betweenness centrality (Brandes, 2001) and eigenvector centrality and the clusters are determined. Finally all other metrics that are mentioned in the previous chapters are calculated dynamically via

NetLogo, being the homophily rate, the perceived value, followers-ratio, heterogeneity (the index of qualitative variation) and the amount of free-rides.

9 NetLogo Model

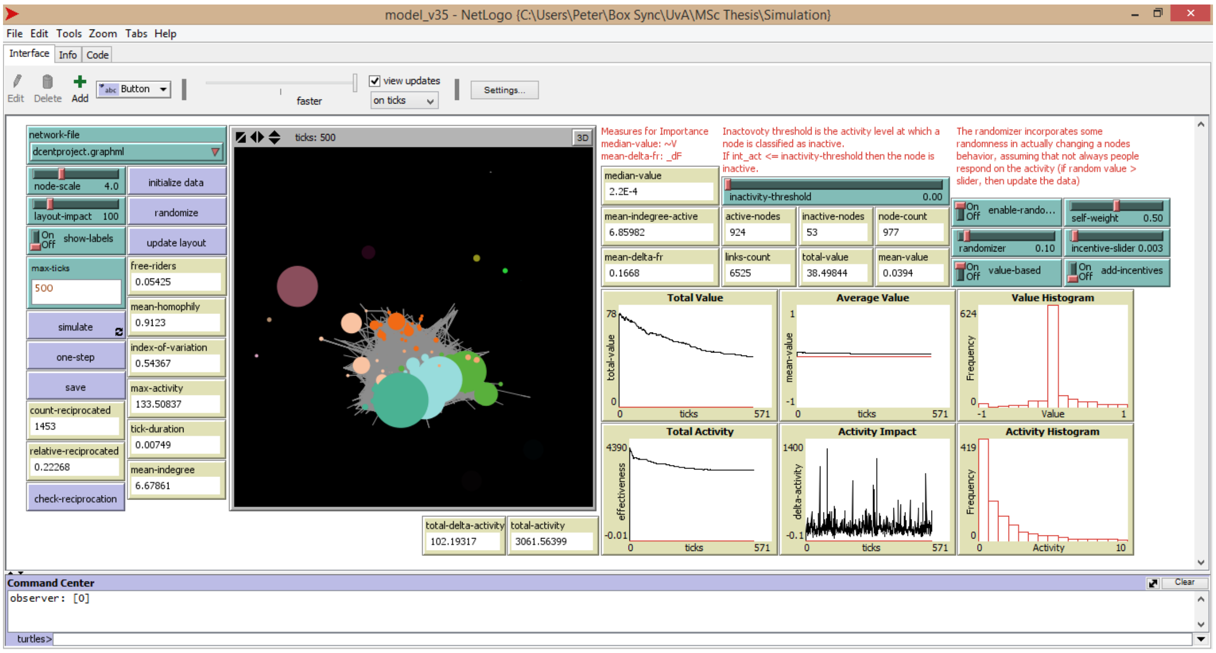


Figure 3: Screenshot Agent-based Simulation

The developed NetLogo model consists of several parts as shown in Figure 3. First the basic operations and presentation options are displayed. The network file to load in the model can be selected, the size for the nodes can be adjusted, which represents the level of activity, and the layout can be reformatted using an energy based algorithm included in NetLogo (see Figure 4). Once the data is initialized some basic properties are calculated and displayed.

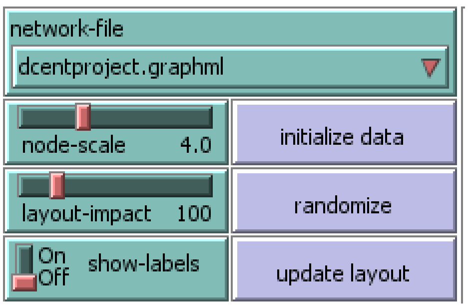


Figure 4: Presentation & Initiation

An option to limit the amount of tics the system processes is implemented, followed by a button to start the simulation or one to execute exactly one step. At any point in time, the simulated network can be saved. Since counting the ties reciprocation level is a slower process, a separate command is implemented to execute this function. The basic properties as mentioned throughout this document are calculated and displayed on the right (see Figure 5). Next to the properties the network graph is visualized in which the size of the node reflects the daily activity (normalized to the selected node-scale).

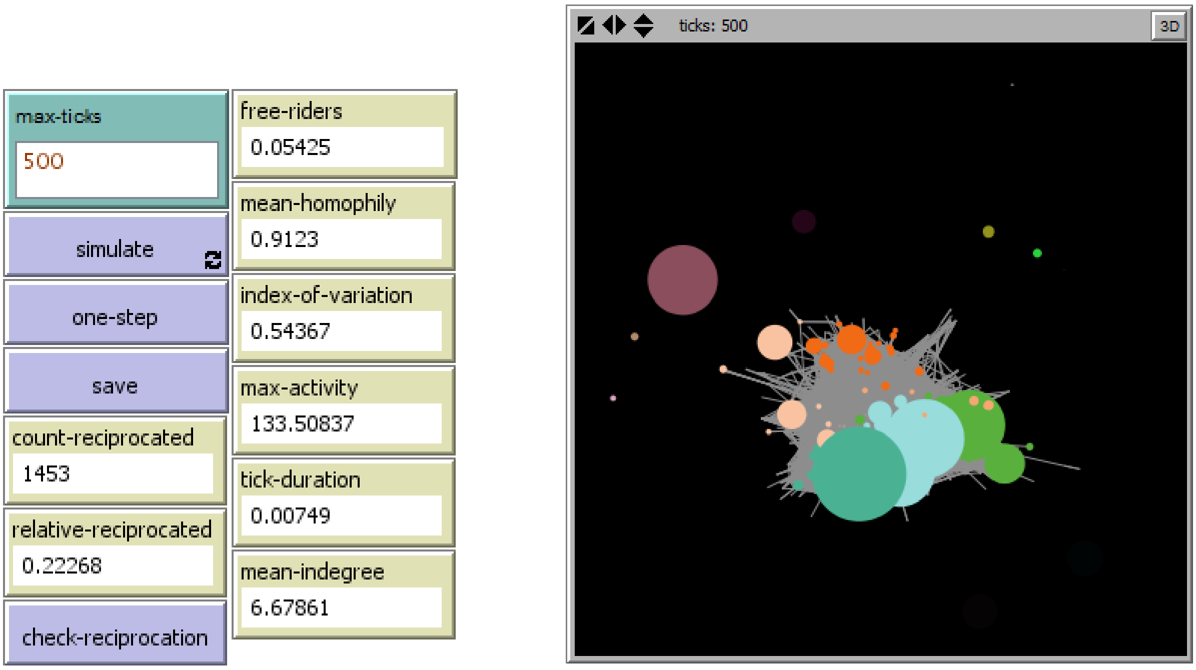


Figure 5: (left) Simulation and Properties and (right) Network Graph Visualization

The different simulation modes as explained in section 7.2 can be selected as well as the randomization properties. To reflect reality slightly better, Figure 6 shows settings in which a level of error can be set in which participants can ignore to update their activity and value level or additional incentives can influence participants to spontaneously contribute to the network. Also a weight factor is implemented to change the amount in which people follow their own beliefs and are susceptible for influence from their friends.

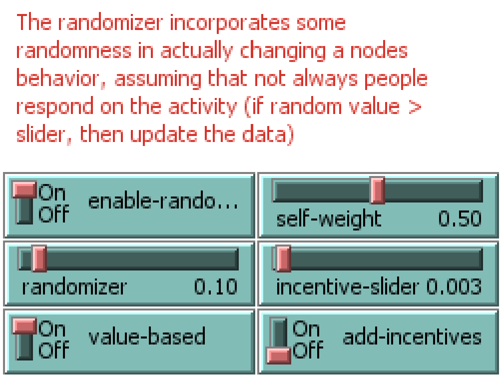


Figure 6: Randomization Options

The dynamic properties that change during the models execution are displayed separately and contain the main outcomes of the formulae specified in the previous chapters. Also an activity threshold is implemented that allows the user to set at what level of activity a node is perceived as inactive (thus being classified as free rider as Figure 7 shows.

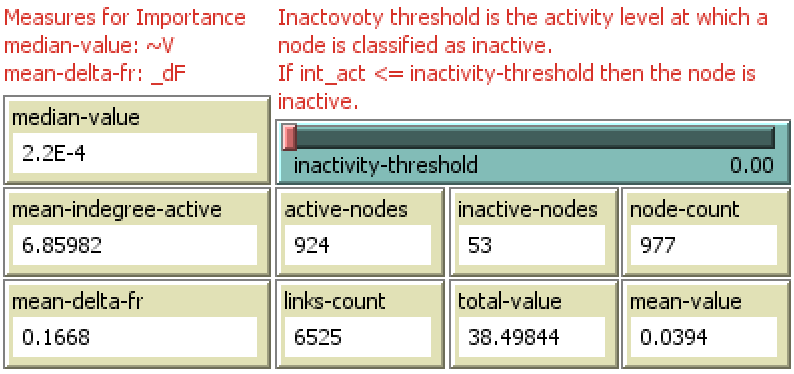


Figure 7: Diverse Dynamic Metrics

To monitor the changes over time and visually see trends in the development, several charts and histograms are designed that show the development of network value and activity as Figure 8 shows.

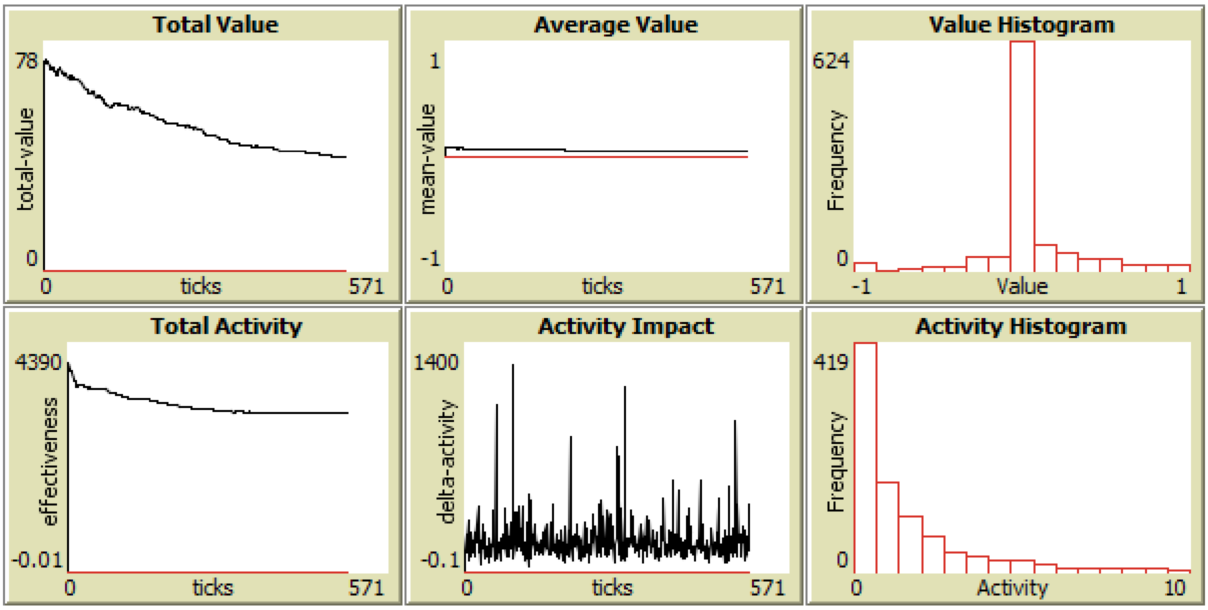


Figure 8: Charts and Histograms

10 Analysis and Results

From the available CAPS funded by the first call, six are selected, due to their differences in topic and representation on Twitter. Table 1 shows the selected networks and the characteristics of their egocentric Twitter network as well as the date of data collection. The collection date is presented in MM/DD/YYYY format and is not part of the parameters for the simulation. The number of nodes and edges are the same in both the real as random network and therefore also the average amount of followers remains equal. The collective value (the average of all perceived values), the initial value (the sum of all perceived values) and the initial activity (total amount of daily tweets) are the input parameters at the time of initiating the simulation (*t=0*), thus their value and the underlying distributions of perceived value and activity are unchanged at the initiation of each simulation. The 3-Months column shows how many ticks the system simulation must process to cover a 3-month time period (as explained in section 7.2).

Table 1: Selected CAPS projects

Project	*Collection Date*	*Count Vertices*	*Count Edges*	*Average Followers*	*Collective Value*	*Initial Value*	*Initial Activity*	*3-Months*
IA4SI	5/3/2016	224	2233	9.96875	0.03384	6.62673	367	3030
My Accessible EU	6/3/2016	315	9621	30.54286	0.02503	9.65019	1240	9000
D-Cent Project	6/6/2016	977	6525	6.67861	0.05314	71.09331	4088	9454
USEMP Project	6/3/2016	66	165	2.67797	0.11348	9.71748	65	1000
WebCOSI	6/3/2016	188	944	5.02128	0.06827	15.96464	580	6000
Comrades Project	4/26/2016	166	1047	8.56024	0.14759	30.18973	1355	18000

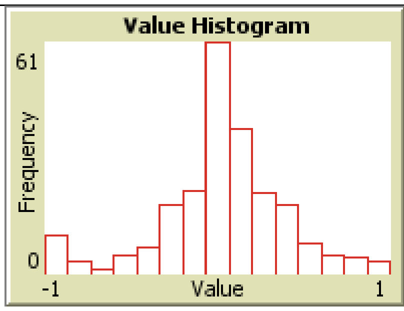
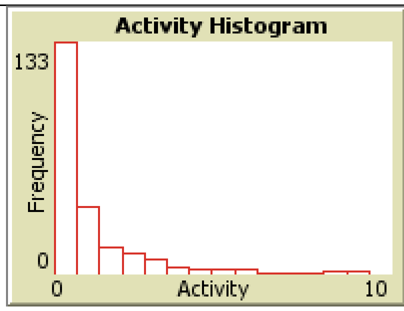
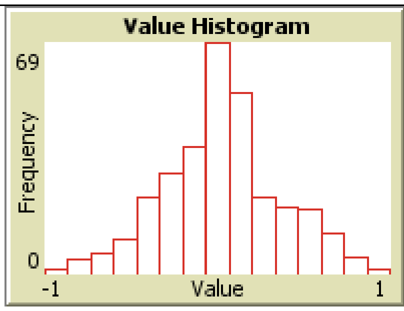
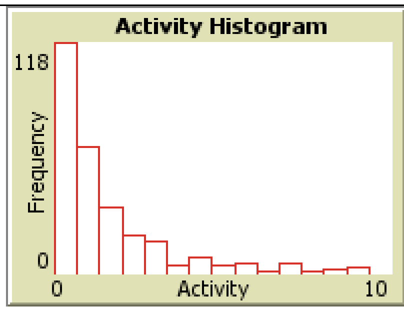
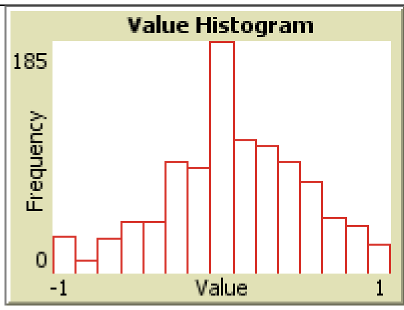
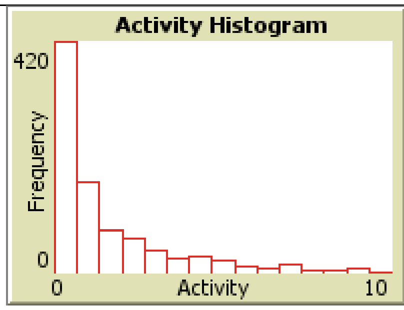
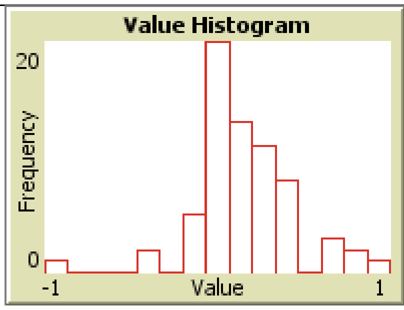
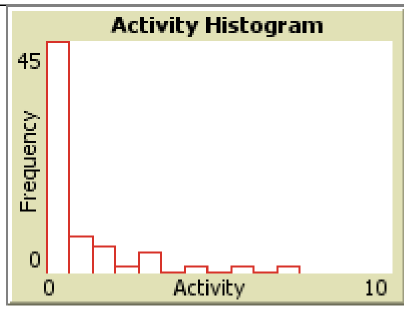
The two robustness indicators, heterogeneity and free-riders, are different between the real and random networks due to the change in ties. The relative amount of free-riders is likely slightly larger in the random network than in the original, because the probability a person being connected to at least one of the other participants is more likely because the links are actively chosen by people instead of just randomly allocated. The level of heterogeneity is expected to be less because the cluster formation in a random network is higly unlikely since there is no preferential attachement involved. Table 2 shows the parameters of the real and random network.

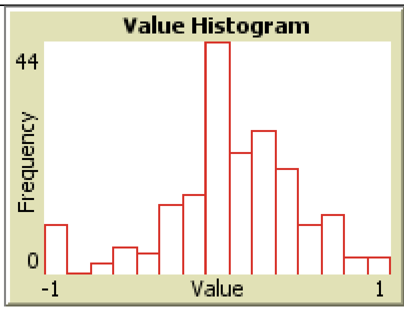
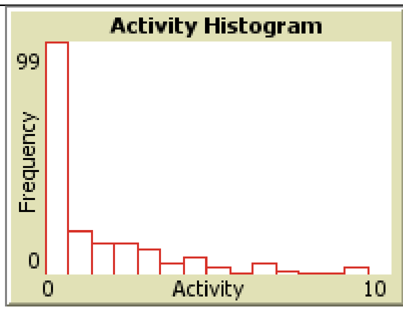
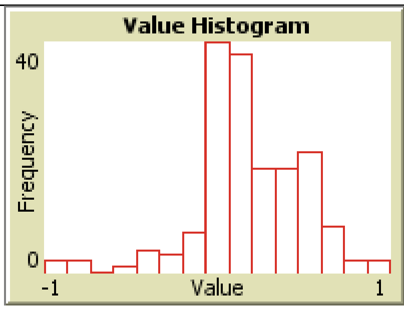
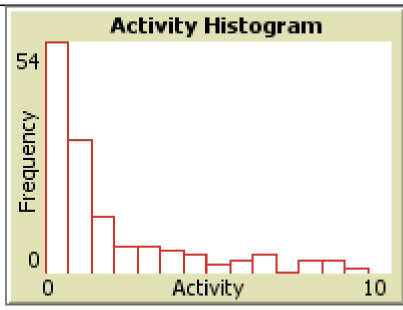
Table 2: Overview Heterogeneity and Free-Riders

	Real Network		**Random Network**	
Project	*Heterogeneity Level*	*Relative Free-Riders*	*Heterogeneity Level*	*Relative Free-Riders*
IA4SI	0.369	9.38%	0.000	10.27%
My Accessible EU	0.099	0.32%	0.000	0.64%
D-Cent Project	0.544	5.73%	0.006	7.57%
USEMP Project	0.615	7.58%	0.473	10.61%
WebCOSI	0.565	6.92%	0.032	7.98%
Comrades Project	0.469	4.82%	0.024	7.23%

Table 3 shows tha histogram of the value distribution and activity distribution for each project that are embedded in the initial network structures which serve as input parameter for the simulation. The distributions indicate that the average participant shows a similar activity level compared with its overall Twitter activity (hence the normal distribution in the value histograms) and that the actual activity in the network is low (following a lognormal distribution). Because the simulation does not use any algorithms that assume any pre-defined distributions testing the normality of each distribution is not performed.

Table 3: Overview Value- and Activity Distribution

Project	*Distribution Initial Value*	*Distribution Initial Activity*
IA4SI		
My Accessible EU		
D-Cent Project		
USEMP Project		

WebCOSI		
Comrades Project		

In order to create a more realistic simulation, a level of randomness is implemented in the algorithm so that not every vertex is activated at each tick and therefore their followers will not update their own activity after each tick. Secondly, when loading the network data into the simulation, NetLogo creates the vertices in a different sequence which also influences the sequence in which the application loops through the vertices to update their data.

Table 4: Simulation Results Real Networks

	Run 1		**Run 2**		**Run 3**		**Run 4**		**Run 5**	
Project	**TV**	**DA**	**TV**	**DA**	**TV**	**DA**	**TV**	**DA**	**TV**	**DA**
IA4SI	-4.59	726	-4.68	726	-4.67	725	-4.77	725	-4.58	726
My Accessible EU	1.37	1201	1.55	1202	1.31	1202	1.42	1202	1.39	1202
D-Cent Project	22.33	2995	22.06	2995	22.32	2998	22.52	3001	21.40	2996
USEMP Project	3.58	43	3.52	43	3.44	43	3.60	43	3.35	43
WebCOSI	1.58	513	0.68	512	0.66	512	0.90	512	1.09	513
Comrades Project	4.90	749	5.01	749	4.94	749	5.02	749	4.81	749

For each network, the simulation is performed 5 times, simulating a timeframe of 3 months of activity. The results are displayed in Table 4 for the real networks and Table 5 for the randomly generated networks. The abbreviations TV and DA stand for Total Value and Daily Activity respectively.

Table 5: Simulation Results Random Networks

Project	Run 1		Run 2		Run 3		Run 4		Run 5	
	TV	DA	TV	DA	TV	DA	TV	DA	TV	DA
IA4SI	0.00	727	0.00	727	0.00	727	0.00	727	0.00	727
My Accessible EU	0.00	1172	0.00	1172	0.00	1172	0.00	1172	0.00	1172
D-Cent Project	-0.36	2768	-0.06	2769	-0.27	2768	0.05	2769	-0.28	2768
USEMP Project	2.64	45	2.26	46	2.96	48	2.48	45	2.82	46
WebCOSI	0.23	466	0.31	467	0.12	465	0.19	467	0.23	467
Comrades Project	-0.11	622	-0.10	622	-0.11	622	-0.11	622	-0.11	622

As a final result a two sample t-test is applied for each CAPS network where the mean total value or mean daily activity of the real network are compared with the ones generated by the random network and are validated at a 95% confidence level (*α=0.05*). The results for the total activity are shown in Table 6.

Table 6: Results 2-Sample t-Test Total Daily Activity

Project	Total Daily Activity				
	Real Network		Random Network		p-Value
	Mean	95% CI	Mean	95% CI	µ1-µ2=0
IA4SI	725.6	[724.9 ; 726.3]	727.0	[727.0 ; 727.0]	0.0004
My Accessible EU	1201.8	[1201.2 ; 1202.4]	1172.0	[1172.0 ; 1172.0]	0.0000
D-Cent Project	2997.0	[2993.8 ; 3000.1]	2768.4	[2767.7 ; 2769.1]	0.0000
USEMP Project	43.0	[43.0 ; 43.0]	46.0	[44.5 ; 47.5]	0.0006
WebCOSI	512.4	[511.7 ; 513.1]	466.4	[465.3 ; 467.5]	0.0000
Comrades Project	749.0	[749.0 ; 749.0]	622.0	[622.0 ; 622.0]	-

For the Comrades Project no –Value could be calculated because no variance has been detected in both simulations. This signifies that the difference is not a detectable probability but a certainty. The results regarding the total value are shown in Table 7.

Table 7: Results 2-Sample t-Test Total Value

Project	Total Value				
	Real Network		Random Network		p-Value
	Mean	95% CI	Mean	95% CI	µ1-µ2=0
IA4SI	-4.658	[-4.757 ; -4.560]	0.000	[0 ; 0]	0.0000

My Accessible EU	1.406	[1.296 ; 1.515]	0.000	[0 ; 0]	0.0000
D-Cent Project	22.127	[21.583 ; 22.670]	-0.183	[-0.398 ; 0.032]	0.0000
USEMP Project	3.498	[3.369 ; 3.627]	2.631	[2.288 ; 2.975]	0.0002
WebCOSI	0.982	[0.515 ; 1.448]	0.217	[0.127 ; 0.307]	0.0021
Comrades Project	4.936	[4.830 ; 5.042]	-0.107	[-0.111 ; -0.103]	0.0000

All p-values, for both the total daily activity and total value, are significantly less than the α-level of 0.05, which leads to the conclusion that the null hypothesis is disproven making it safe to assume that the structure of the network plays an important role in the sustainability of the CAPS network. The extent in which the predicted total value and daily activity are correctly estimated remains an open question until the first CAPS cease to exist and the experiment is repeated several times.

Finally using linear regression, the heterogeneity indicator and the relative amount of free-riders are used to explain the total daily activity or total value. The first model is defined as:

$$\bar{a} = \beta_0 + \beta_1 \bar{a}_{t=0} + \beta_2 I + \beta_3 f + \varepsilon$$

where $\bar{a}$ is the average daily activity, let $\bar{a}_{t=0}$ be the initial daily activity, I the heterogeneity level and f the relative amount of free-riders and all β-values are the coefficients. Performing this regression results in the coefficients shown in Table 8 (where the significance is indicated with *** for α <= 0.01, ** for α <= 0.05 and * for α <= 0.10).

Table 8: Regression Results on Daily Activity

Variable	**Coefficient**	**Standard Error**
Initial Activity ($\bar{a}_{t=0}$)	$\beta_1 =$ 0.6478122 ***	0.062873
Heterogeneity (I)	$\beta_2 =$ -142.3333	336.0397
Free-Riders (f)	$\beta_3 =$ -988.0636	2680.935
Constant	$\beta_0 =$ 274.6965	233.1262

Although the adjusted R-squared value shows that approximately 91% of all variance can be explained by the independent variables the only significant one is the initial activity, thus the heterogeneity and free-riders do not explain the sustainability.

The second model, where the relation is based on the total value, is defined as:

$$V = \beta_0 + \beta_1 V_{t=0} + \beta_2 I + \beta_3 f + \varepsilon$$

where V is the total value, let $V_{t=0}$ be the initial total value, I the heterogeneity level and f the relative amount of free-riders and all β-values are the coefficients. Performing this regression results in the coefficients shown in Table 8 (where the significance is indicated with *** for α <= 0.01, ** for α <= 0.05 and * for α <= 0.10).

Variable	Coefficient	Standard Error
Initial Value ($V_{t=0}$)	β_1 = 0.165808 **	0.0618747
Heterogeneity (I)	β_2 = 12.49575 *	5.741788
Free-Riders (f)	β_3 = -45.31883	44.79404
Constant	β_0 = -1.730391	3.675968

The adjusted R-squared value, 0.4670, shows that approximately 47% of all variance can be explained by the independent variables (the overall level of fit). The dataset is small, 12 observations in total, so whether the R^2 value shows a significant fit is arbitrary. The model shows that heterogeneity has a positive and somewhat significant impact on the value, thus can explain sustainability to a certain extent.

11 Conclusion and Discussion

This study answers the question to what extent the structure of a network determines the sustainability of an egocentric network. The developed conceptual framework fills a gap in extant literature by explaining sustainability of a network in terms of its importance (section 3), effectiveness (section 4) and robustness (section 5) instead of using a general utility function i.e. a cost-benefit analysis. This is an important difference because it avoids a complex discussion on determining costs and benefits, which are experienced differently for each network member, and uses the outcome of any such function in terms of activity instead. Another contribution to the current body of literature is that two different sources of information are combined. The structural information about the friendships in a network is joined with communication patterns of the network members. The newly created dataset is used in an agent-based simulation, providing insight in the expected activity development and therefore sustainability of the network.

Based on this decomposition of sustainability, measures and formulas are assigned to assess each segment. For *importance*, the ability of the network to effect reality, both the perceived and collective value are defined in terms of internal and external activity, assuming that the activity correlates with the perceived value. Further, the amount of followers relative to all one's connections is calculated for within the network as in general, which signifies whether the user has more than average followers making the network a more valuable instrument for the user. To assess the *effectiveness*, the ability of the network structure to affect and transfer action, a discrete-time model is developed that predicts a user's influence. It uses a homophily indicator for the clusters to determine the weight of each tie and takes the user's previous activity in combination with the user's exposure to others to determine the future activity. Finally the *robustness* of the network is assessed, i.e. the ability of the network structure to overcome disturbances, in terms of heterogeneity and free-riders.

The empirical data of this study shows that the structure of an egocentric network is of key importance to the sustainability of the network, because it affects the perceived value (i.e. *importance*) as well as the total activity (i.e. *effectiveness*). The study demonstrates that when no data is available (i.e. no CAPS is dissolved yet) it is possible to create a simulation based on a strong theoretical review. The developed agent based model calculates all the measures mentioned, providing an a priori insight on the sustainability of the network after the CAPS is dissolved. This empowers the CAPS to create a more sustainable structure during its lifetime, a condition that can be made during the provisioning of future grants to ratify a societal or communal return-on-investment besides the existing knowledge based return-on-investment.

The *robustness* is tested empirically using a linear regression approach, because it does not change during the simulation. While new entry of participants and existing participants leaving the network can be simulated, it changes the structure of the network, thus the output of the simulation will not predict the importance of the network structure alone, but will be contaminated, reducing the validity of the results. Besides the initial daily activity and the initial total value, only the heterogeneity level is found to be significant in influencing the sustainability of the network assessed from the value based approach.

Although the conceptual framework this study develops serves as a valuable tool for assessing the sustainability of a network, as mentioned, detailed empirical data is not yet present in the realm of CAPS. A valuable contribution to this topic would be to test the measures and algorithms in several scenarios or case studies in which the sustainability is known (either after a CAPS is dissolved or comparing similar structures on Twitter in which the ego is known to be removed from the network. Another important area of improvement is to study the level of influence in more detail, by combining e.g. psychological knowledge and empirical data capturing this knowledge on people involved in a social movement. Also research that identifies what kind of patters result in an improved sustainability will contribute greatly. The agent based model in this study provides several variables that can be changed to assess the reaction on the network, but further research could identify what the value of these variables should ideally be to reflect reality.

By changing the given parameters of randomness and self-weight, the distribution of the total value and collective action could be spread out, i.e. the domain of the 95% confidence interval could grow, making the t-tests less evident, potentially disproving the hypothesis this research focuses on. A final issue for improvement is that this study takes the egocentric networks as static, while in real life they are dynamic. Making the model dynamic and allowing the simulation to have people joining and leaving the network, and potentially simulating different external forces that influence the network, will increase the usability of the model to date.

Finally from a practical perspective, the conceptual framework helps CAPS initiatives to assess the sustainability of their network, supported by the agent-based model to visualize the process. The model emphasizes researching the effectiveness of network structure and enables testing possible strategies for platform development. The CAPS can alter the network structure to assess the consequences or assess to what extent additional incentives are required to increase the network's value. The outcome of this research also provides justification for the European Union's funding supporting the CAPS by enabling them to improve continuity of awareness due to the sustainability of the created network. As this study indicates, social innovation is not a product of an organization which is (financially) supported by the European Commission (EC). Social innovation requires a blended approach that empowers and facilitates a community in solving the problems at hand, a role that both the CAPS and the EC can provide more extensively when it knows how to improve its network sustainability.

12 References

Agresti, A., & Agresti, B. F. (1978). Statistical Analysis of Qualitative Variation. In *Sociological Methodology* (Vol. 9, pp. 204–237).

Aizstrauta, D., Ginters, E., & Eroles, M.-A. P. (2015). Applying Theory of Diffusion of Innovations to Evaluate Technology Acceptance and Sustainability. *Procedia Computer Science*, *43*(0), 69–77. http://doi.org/http://dx.doi.org/10.1016/j.procs.2014.12.010

Alhabash, S., & McAlister, a. R. (2014). Redefining virality in less broad strokes: Predicting viral behavioral intentions from motivations and uses of Facebook and Twitter. *New Media & Society*, 1–23. http://doi.org/10.1177/1461444814523726

Anand, S., & Sen, A. (2000). Human development and economic sustainability. *World Development*, *28*(12), 2029–2049. http://doi.org/10.1016/S0305-750X(00)00071-1

Arniani, M., Badii, A., Liddo, A. De, Georgi, S., Passani, A., Piccolo, L. S. G., & Teli, M. (2014). Collective Awareness Platform for Sustainability and Social Innovation: An Introduction. Retrieved from http://booksprints-for-ict-research.eu/wp-content/uploads/2014/07/BS5-CAPS-FIN-003.pdf

Bajpai, K. (2011). A Framework for Analyzing Collective Action Events on Twitter. *The 8th International ISCRAM Conference*, (May).

Barabasi, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. http://doi.org/DOI 10.1126/science.286.5439.509

Barabási, A.-L., & Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, (May), 50–59.

Baraglia, R., Dazzi, P., Mordacchini, M., & Ricci, L. (2013). A peer-to-peer recommender system for self-emerging user communities based on gossip overlays. *Journal of Computer and System Sciences*, *79*(2), 291–308. http://doi.org/10.1016/j.jcss.2012.05.011

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10008*(10), 6. http://doi.org/10.1088/1742-5468/2008/10/P10008

Borodin, A., Filmus, Y., & Oren, J. (2010). Threshold Models for Competitive Influence in Social Networks. *Science*, 1–15.

Boudin, L., Salvarani, F., Boudin, L., & Salvarani, F. (2015). Opinion dynamics : kinetic modelling with mass media , application to the Scottish independence referendum. *Physica A*, *444*, 448–457. http://doi.org/10.1016/j.physa.2015.10.014

Brandes, U. (2001). A faster algorithm for betweenness centrality*. *The Journal of Mathematical Sociology*, *25*(2), 163–177. http://doi.org/10.1080/0022250X.2001.9990249

Centola, D. (2013). A Simple Model of Stability in Critical Mass Dynamics. *Journal of Statistical Physics*, *151*(1-2), 238–253. http://doi.org/10.1007/s10955-012-0679-3

Chang, C. (2015). Self-construal and Facebook activities: Exploring differences in social interaction orientation. *Computers in Human Behavior*, *53*, 91–101. http://doi.org/10.1016/j.chb.2015.06.049

Christensen, C. (2011). Twitter Revolutions? Addressing Social Media and Dissent. *The Communication Review*, *14*(3), 155–157. http://doi.org/10.1080/10714421.2011.597235

Cohen, R., Erez, K., Ben-Avraham, D., & Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical Review Letters*, *85*(21), 4626–4628. http://doi.org/10.1103/PhysRevLett.85.4626

Couture-Beil, A. (2014). rjson: JSON for R. Retrieved from https://cran.r-project.org/package=rjson

Davis, F. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of ManMachine Studies*. http://doi.org/10.1006/imms.1993.1022

DeGroot, M. H. (1974). Reaching a Consensus. *Journal of American Statistical Association*, *69*(345), 118–121.

Diaz-aviles, E., Lam, H. T., Pinelli, F., Braghin, S., Gkoufas, Y., Berlingerio, M., & Calabrese, F. (2014). Predicting User Engagement in Twitter with Collaborative Ranking. *Proceedings of the 2014 Recommender Systems Challenge on - RecSysChallenge '14*, 41–46. http://doi.org/10.1145/2668067.2668072

Doong, S. H. (2016). Predicting Twitter Hashtags Popularity Level. *Proceedings of 49th Hawaii International Conference on System Sciences*, 1959–1968. http://doi.org/10.1109/HICSS.2016.247

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493. http://doi.org/10.1016/0047-2484(92)90081-J

Fan, L., & Suh, Y. H. (2014). Why do users switch to a disruptive technology? An empirical study based on expectation-disconfirmation theory. *Information and Management*, *51*(2), 240–248. http://doi.org/10.1016/j.im.2013.12.004

Geddes, C. (2011). Achieving critical mass in social networks. *Journal of Database Marketing & Customer Strategy Management*, *18*(2), 123–128. http://doi.org/10.1057/dbm.2011.14

Gentry, J. (2015). twitteR: R Based Twitter Client. Retrieved from https://cran.r-project.org/package=twitteR

Go, E., & You, K. H. (2016). But not all social media are the same: Analyzing organizations' social media usage patterns. *Telematics and Informatics*, *33*(1), 176–186. http://doi.org/10.1016/j.tele.2015.06.016

Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on twitter networks: Validation of Dunbar's number. *PLoS ONE*, *6*(8). http://doi.org/10.1371/journal.pone.0022656

Gruzd, a., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an imagined community. *American Behavioral Scientist*, *55*(10), 1294–1318. http://doi.org/10.1177/0002764211409378

Harris, G. P. (2007). *Seeking Sustainability in an Age of Complexity*. Cambridge: Cambridge University Press.

Hollenbaugh, E. E., & Ferris, A. L. (2014). Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. *Computers in Human Behavior*, *30*, 50–58. http://doi.org/10.1016/j.chb.2013.07.055

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jalili, M. (2012). Social power and opinion formation in complex networks. *Physica A*, *392*(4), 1–8. http://doi.org/10.1016/j.physa.2012.10.013

Kim, H., & Bearman, P. S. (1997). The Structure and Dynamics of Movement Participation. *American Sociological Review*, *62*(1), 70–93. http://doi.org/10.2307/2657453

Kindler, A., Solomon, S., & Stauffer, D. (2013). Peer-to-peer and mass communication effect on opinion shifts. *Physica A: Statistical Mechanics and Its Applications*, *392*(4), 785–796. http://doi.org/10.1016/j.physa.2012.10.038

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (Second Edi). Thousand Oaks, CA: SAGE Publications.

Lee, Y., Lee, J., & Hwang, Y. (2015). Relating motivation to information and communication technology acceptance: Self-determination theory perspective. *Computers in Human Behavior*, *51*, 418–428. http://doi.org/10.1016/j.chb.2015.05.021

Li, C. Y. (2013). Persuasive messages on information system acceptance: A theoretical extension of elaboration likelihood model and social influence theory. *Computers in Human Behavior*, *29*(1), 264–275. http://doi.org/10.1016/j.chb.2012.09.003

Lu, Y., Zhang, P., Cao, Y., Hu, Y., & Guo, L. (2014). On the frequency distribution of retweets. *Procedia Computer Science*, *31*, 747–753. http://doi.org/10.1016/j.procs.2014.05.323

Ma, Z., Sun, A., & Cong, G. (2013). On Predicting the Popularity of Newly Emerging Hashtags in Twotter. *Journal of the American Society for Information Science and Technology*, *64*(7), 1399–1410. http://doi.org/10.1002/asi

Marwell, G., Oliver, P., & Prahl, R. (1988). Social Networks and Collective Action: A Theory of Critical Mass III.pdf. *American Journal of Sociology*, *94*(3), 502–534.

Mcpherson, M., Smith-lovin, L., & Cook, J. M. (2016). BIRDS OF A FEATHER : Homophily in Social Networks, *27*(2001), 415–444.

Men, L. R., & Tsai, W.-H. S. (2015). Infusing social media with humanity: Corporate character, public engagement, and relational outcomes. *Public Relations Review*, *41*(3), 395–403. http://doi.org/10.1016/j.pubrev.2015.02.005

Morozov, E. (2009). Iran: Downside to the "Twitter Revolution." *Dissent*, *56*(4), 10–14. http://doi.org/10.1353/dss.0.0092

Mouakket, S. (2015). Factors influencing continuance intention to use social network sites: The Facebook case. *Computers in Human Behavior*, *53*, 102–110. http://doi.org/10.1016/j.chb.2015.06.045

Mrvar, A., & Batagelj, V. (2016). Pajek / Pajek XXL. Ljubljana: University of Ljubljana. Retrieved from http://mrvar.fdv.uni-lj.si/pajek/

Newman, M. E. J. (2004). Detecting community structure in networks. *European Physical Journal B*, *38*(2), 321–330. http://doi.org/10.1140/epjb/e2004-00124-y

Nguyen, H., & Zheng, R. (2014). A data-driven study of influences in Twitter communities. *2014 IEEE International Conference on Communications, ICC 2014*, 3938–3944. http://doi.org/10.1109/ICC.2014.6883936

Nik-Bakht, M., & El-Diraby, T. E. (2015). Sus-tweet-ability: Exposing public community's perspective on sustainability of urban infrastructure through online social media. *International Journal of Human-Computer Studies*, 1–19. http://doi.org/10.1016/j.ijhcs.2015.11.002

Oh, H. J., Ozkaya, E., & LaRose, R. (2014). How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. *Computers in Human Behavior*, *30*, 69–78. http://doi.org/10.1016/j.chb.2013.07.053

Oliver, P., Marwell, G., & Teixeira, R. (1985). A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action. *American Journal of Sociology*, *91*(3), 522. http://doi.org/10.1086/228313

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

Purves, D. (2008). *Principles of cognitive neuroscience*. Sunderland, MA: Sinauer Associates Inc.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rauniar, R., Rawski, G., Yang, J., & Johnson, B. (2014). Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. *Journal of Enterprise Information Management*, *27*(1), 6–30. http://doi.org/10.1108/JEIM-04-2012-0011

Robson, C. (2011). *Real World Research 3e*. *Wiley*. Wiley. Retrieved from http://www.amazon.co.uk/Real-World-Research-Colin-Robson/dp/1405182407

Rogers, E. M. (1995). *Diffusion of innovations*. *Newyork Free Press*. http://doi.org/citeulike-article-id:126680

Romm, C., Pliskin, N., & Clarke, R. (1997). Virtual communities and society: Toward an integrative three phase model. *International Journal of Information Management*, *17*(4), 261–270. http://doi.org/10.1016/S0268-4012(97)00004-2

Sayama, H. (2015). *Introduction to the modeling and analysis of complex systems*. Geneseo: Open SUNY Textbooks.

Segerberg, A., & Bennett, W. L. (2011). Social Media and the Organization of Collective Action: Using Twitter to Explore the Ecologies of Two Climate Change Protests. *The Communication Review*, *14*(3), 197–215. http://doi.org/10.1080/10714421.2011.597250

Shapiro, C., & Varian, H. R. (1999). *Information rules: A Strategic Guide to the Networked Economy*. *ACM SIGMOD Record* (Vol. 32). Boston, MA: Harvard Business School Press. http://doi.org/10.1145/776985.776997

Smith, B. G. ., & Gallicano, T. D. . (2015). Terms of engagement: Analyzing public engagement with organizations through social media. *Computers in Human Behavior*, *53*, 82–90. http://doi.org/10.1016/j.chb.2015.05.060

Social Media Research Foundation. (2014). NodeXL; Network Overview, Discovery and Exploration for Excel. Retrieved from http://nodexl.codeplex.com/Wikipage?ProjectName=nodexl\nhttp://nodexl.codeplex.com/

Venkatesh, V., & Davis, F. (2000). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. http://doi.org/10.1287/mnsc.46.2.186.11926

Vogler, R. (2014). RTwitterAPI: Facilitates GET requests to Twitter API. Retrieved from https://github.com/joyofdata/RTwitterAPI

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629–635. http://doi.org/10.1016/j.joi.2010.07.002

Wang, B., Tang, H., Guo, C., & Xiu, Z. (2005). Entropy Optimization of Scale-Free Networks Robustness to Random Failures, (February 2008). http://doi.org/10.1016/j.physa.2005.08.025

Wang, G., Liu, Y., Li, J., Tang, X., & Wang, H. (2015). Superedge coupling algorithm and its application in coupling mechanism analysis of online public opinion supernetwork. *Expert Systems with Applications*, *42*(5), 2808–2823. http://doi.org/10.1016/j.eswa.2014.11.026

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. *Chemistry & …*, 825. http://doi.org/10.1017/CBO9780511815478

Wilensky, U. (1999). NetLogo 5.3.1. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Retrieved from http://ccl.northwestern.edu/netlogo/

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Base Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge: The MIT Press.

Xu, W. W., Sang, Y., Blasiola, S., & Park, H. W. (2014). Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election. *American Behavioral Scientist*, 0002764214527091–. http://doi.org/10.1177/0002764214527091

Yu, X., & Xie, J. Q. (2014). Modeling Mutual Influence Between Social Actions and Social Ties. *The 25th International Conference on Computational Linguistics*, 848–859.

Zhang, P., Wang, X., & Li, B. (2013). On predicting Twitter trend: factors and models. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1427–1429. http://doi.org/10.1145/2492517.2492576

Appendix A Source Code for Simulation

```
extensions [ nw ]
globals [ total-delta-activity max-size cat-list iqv first-max-activity ]
turtles-own [ x y vertex ext_act int_act value ext_fr int_fr delta_fr
indegree outdegree degree modularity_class eigencentrality strongcompnum
clustering homophily inactive tick-probability ]
links-own [ homophily-weight reciprocated ]
```

```
; ----- BUTTONS -----

to init
  clear-all
  reset-ticks
  nw:set-context turtles links
  nw:load-graphml network-file [
    set shape "circle"
    set size 1
    set label ""
    set color modularity_class + 25
  ]

  ask turtles [
    set homophily 99
  ]

  set-coordinates
  normalize-ticks
  calculate-properties
  calculate-homophily
  calculate-weight
  calculate-value
  calculate-iqv
  set max-size [int_act] of max-one-of turtles [int_act]
  set-size
  set first-max-activity max-activity
end

to layout
  ifelse show-labels
  [ ask turtles [ set label vertex ] ]
  [ ask turtles [ set label "" ] ]
  repeat layout-impact [ layout-spring turtles links 1 5 1 ] ;; lays the
nodes in a triangle
end

to randomize
  ask turtles [
    setxy random-xcor random-ycor
  ]
end

to simulate
  ; Calculate the new activity of each turtle

  ifelse value-based [
    simulate-value
  ]
  [
    simulate-activity
    calculate-value
  ]

  set-size
  if (ticks = max-ticks) [ stop ]
  tick
end

to one-step
  simulate
```

```
  tick
end

to save
  let len-filename length network-file - 8
  let filename substring network-file 0 len-filename
  nw:save-graphml (word filename "-" ticks ".graphml" )
end

; ----- TOOLS -----

to set-coordinates
  ;Determine max screen width and height
  let screen-width abs(min-pxcor) + max-pxcor
  let screen-height abs(min-pycor) + max-pycor

  ;Get smallest and largest coordinates + 10% or border
  let min-x [x] of min-one-of turtles [x] * 1.1
  let max-x [x] of max-one-of turtles [x] * 1.1
  let min-y [y] of min-one-of turtles [y] * 1.1
  let max-y [y] of max-one-of turtles [y] * 1.1

  let canvas-width abs(min-x) + max-x
  let canvas-height abs(min-y) + max-y

  ask turtles [
    let node-x (x + abs(min-x)) * (screen-width / canvas-width) + min-pxcor
    let node-y (y + abs(min-y)) * (screen-height / canvas-height) + min-
pycor
    setxy node-x node-y
  ]
end

to set-size
  ask turtles [
    ifelse (int_act > 0) [ set size ( (int_act * node-scale) / (max-size) )
] [ set size 0 ]
  ]
end

to calculate-properties
  ask turtles [
    if (degree > 0) [
      set int_fr (indegree / degree)
      set delta_fr int_fr - ext_fr
      ifelse ( int_act <= inactivity-threshold ) [ set inactive 1 ] [ set
inactive 0 ]
    ]
  ]
end

to calculate-homophily

  ; Count number of nodes with not homophily (set to 99 by default)
  ; If this count is > 0
  ;   Take one node with no homophily assigned
  ;   Ask all nodes from the same cluster
  ;   Ask all outgoing ties of these nodes
  ;   Count the tie as internal or external regarding the end2
  ;   Calculate the homophily ratio
  ;   Assign this ratio to all nodes with the same cluster
  ;   Execute procedure once more
```

```
  ; Otherwise stop

  if (count turtles with [ homophily = 99 ] > 0) [
    let internal-links 0
    let external-links 0
    ask one-of turtles with [ homophily = 99 ] [
      let current-cluster strongcompnum
      ask turtles with [ strongcompnum = current-cluster ] [
        ask my-out-links [
          ifelse ( [strongcompnum] of end2 = current-cluster ) [
            set internal-links internal-links + 1
          ]
          [
            set external-links external-links + 1
          ]
        ]
      ]
      ask turtles with [ strongcompnum = current-cluster ]  [
        ifelse (internal-links + external-links > 0) [
          set homophily internal-links / ( external-links + internal-links
)
        ]
        [
          set homophily 1
        ]
      ]
    ]
    calculate-homophily
  ]
end

to calculate-weight

  ; For each node assign a weight to the outgoing ties
  ; The weight is 1/d of the node itself * the homophily for internal links
  ; and 1-homophily for external links, where d is the total degrees of the
node

  ask turtles [
    ask my-out-links [
      ifelse ( [strongcompnum] of end1 = [strongcompnum] of end2 )
      [
        ifelse ( [degree] of end1 > 0 ) [
          set homophily-weight ( 1 / ([degree] of end1 ) ) * [homophily] of
end1
        ] [ set homophily-weight 1 ]
      ]
      [
        ifelse ( [degree] of end1 > 0 ) [
          set homophily-weight ( 1 / ([degree] of end1 ) ) * ( 1 -
[homophily] of end1 )
        ] [ set homophily-weight 1 ]
      ]
    ]
  ]
end

to provide-incentive
  if (add-incentives) [
    if ( ( random 100 / 100 < incentive-slider ) or ( not enable-randomizer
) ) [
      ask one-of turtles [
```

```
        set int_act ( random ( first-max-activity * 100 ) / 100 )
      ]
    ]
  ]
end

to simulate-activity

  ; For each node
  ; Calculate the sum of all (activity * weight) of its neighbors
  ; Add the nodes own activity (average is already included in homophily-
weight
  ; Upate the value for the nodes

  ; Create a random number, for each node with tick-probability >= random
number
  ; Get their followers (in-inks) and update them with the int_act
  let new-act 0
  set total-delta-activity 0
  ask turtles with [ tick-probability >= ( random 1000 / 1000 ) ] [
    ask my-in-links [
      set new-act ( homophily-weight * ( [int_act] of end2 - [int_act] of
end1 ) )
      if ( ( random 100 / 100 < randomizer ) or ( not enable-randomizer ) )
[
        ask end1 [
          set int_act ( self-weight * int_act ) + ( 1 - self-weight ) *
new-act
          ;ifelse ( new-act < inactivity-threshold ) [ set inactive 1 ] [
set inactive 0 ]
          ifelse ( int_act <= inactivity-threshold ) [ set inactive 1 ] [
set inactive 0 ]
        ]
      ]
    ]
    set total-delta-activity ( total-delta-activity + new-act )
  ]
  provide-incentive

  ; Potentially adjust with centrality (importance)
  ; by taking the highes centrality measure of all of its neighbors & self
  ; and dividing each centrality by this maximum value ( so normalize to
[0, 1] )
  ; and then mulitply the transferrable activity with this factor

end

to simulate-value

  ; For each node
  ; Calculate the sum of all (value * weight) of its neighbors
  ; Add the nodes own value (average is already included in homophily-
weight

  ; Create a random number, for each node with tick-probability >= random
number
  ; Get their followers (in-inks) and update them with the int_act
  let new-value 0
  let new-act 0
  set total-delta-activity 0
  ask turtles with [ tick-probability >= ( random 1000 / 1000 ) ] [
    ask my-in-links [
```

```
      set new-value ( 1 - self-weight ) * ( homophily-weight * [value] of
end2 ) + ( self-weight * [value] of end1 )
      if ( ( random 100 / 100 < randomizer ) or ( not enable-randomizer ) )
[
        ask end1 [
          ;ifelse ( new-value < value - inactivity-threshold ) and ( new-
value > value + inactivity-threshold ) [ set inactive 0 ] [ set inactive 1
]
          ifelse ( new-value != 1 ) [
            set new-act ( ( ( - 1 * new-value ) - 1 ) * ext_act ) / ( new-
value - 1 )
          ] [ set new-act int_act ]
          set value new-value
          let delta-act new-act - int_act
          set total-delta-activity ( total-delta-activity + delta-act )
          set int_act new-act
          ifelse ( int_act <= inactivity-threshold ) [ set inactive 1 ] [
set inactive 0 ]
        ]
      ]
    ]
    set total-delta-activity ( total-delta-activity + new-act )
  ]
  provide-incentive

end

to calculate-value
  ask turtles [
    ifelse (int_act + ext_act > 0)
    [ set value (int_act - ext_act) / (int_act + ext_act) ]
    [ set value 0 ]
  ]
end

to check-reciprocation
  ask links with [ reciprocated = 0 ] [
    ifelse count links with [ ( end1 = [end2] of myself ) and ( end2 =
[end1] of myself ) ] > 0
    [ set reciprocated 1]
    [ set reciprocated -1]
  ]
end

to normalize-ticks

  ; Normalizing the ticks is required for proper time simulation, where the
largest tweeter tweets every tick
  ask turtles [
    set tick-probability int_act / max-activity
  ]
end

to calculate-iqv

  ; Calculating the IQV results in an indication for potential collective
action
  ; The higher the score, the higher the heterogeneity, score 1 means 50%
chance eternally connected

  set cat-list [ ]
  ask turtles [
```

```
    set cat-list lput strongcompnum cat-list
  ]
  set cat-list remove-duplicates cat-list
  let cat-sum 0
  foreach cat-list [
    let cat-prop count turtles with [ strongcompnum = ? ] / count turtles
    set cat-sum cat-sum + cat-prop ^ 2
  ]
  ifelse ( length cat-list = 0 ) or ( length cat-list = 1 )
  [ set iqv 0 ]
  [ set iqv ( 1 - cat-sum ) / ( 1 - ( 1 / length cat-list ) ) ]
end

; ----- REPORTERS -----

to-report median-value
    report median [value] of turtles
end

to-report mean-delta-fr
  report mean [delta_fr] of turtles
end

to-report mean-indegree-active
  report mean [indegree] of turtles with [int_act > 0]
end

to-report node-count
  report count turtles
end

to-report active-nodes
  ;report count turtles with [int_act > inactivity-threshold]
  report count turtles with [inactive = 0]
end

to-report inactive-nodes
  ;report count turtles with [int_act <= inactivity-threshold]
  report count turtles with [inactive = 1]
end

to-report links-count
  report count links
end

to-report free-riders
  report inactive-nodes / node-count
end

to-report mean-homophily
  report mean [homophily] of turtles
end

to-report total-activity
  report sum [int_act] of turtles
end

to-report total-value
  report sum [value] of turtles with [(value >= 0) or (value <= 0)]
end

to-report mean-value
```

```
  ifelse (ticks > 0) [
    ;report mean [value] of turtles with [(value >= 0) or (value <= 0)]
    report mean [value] of turtles
  ] [ report 0 ]
end

to-report mean-indegree
  let total-in-links 0
  ask turtles [
    set total-in-links total-in-links + count my-in-links
  ]
  report total-in-links / node-count
end

to-report total-delta
  report total-delta-activity
end

to-report count-reciprocated
  report ( count links with [ reciprocated = 1 ] ) / 2
end

to-report relative-reciprocated
  report ( ( count links with [ reciprocated = 1 ] ) / 2 ) / links-count
end

to-report turtle-values
  ask turtles [
    report value
  ]
end

to-report max-activity
  report max [int_act] of turtles
end

to-report index-of-variation
  report iqv
end

to-report tick-duration
  report 1 / max-activity
end
```