2009 International Conference on Computational Science and Engineering

Clustering Signature in Complex Social Networks

Ueli Peter and Tomas Hruz
Theoretical Computer Science, ETH Zurich, Switzerland
e-mail: tomas.hruz@inf.ethz.ch

***Abstract*—An important aspect in social computing is the structure of social networks, which build the underlying substrate for the exchange of information. With the growing importance of microblogging networks like Twitter a new class of directed social networks appeared. Recently, Ahnert and Fink [1] showed that some classes of directed networks are cleanly separated in the space of the clustering signature. In this paper, we study the structural dynamics which defines the clustering signature in scale free equilibrium networks. Moreover, we also study the hierarchical features of the network topology which lead to a deeper understanding of the clustering signature.**

I. Introduction

Social interaction between people in the age of increased depth, breath and speed of communication and computing facilities needs new research into aspects which were of limited interest until recently. For example, there are cases, where political decisions are considerably influenced by the possibility to quickly disseminate information and organize large groups of people through technological platforms (see [2] and references therewith). It is well known that a social network has the ability to achieve interactively - "online" a collective decision and knowledge, that can be considered from the system point of view as a solution to a certain computing problem [3]. This information processing occurs through an underlying network of interpersonal connections. When people want to share information and think collectively about some problem, apart from using cultural infrastructure - books, scientific publications and arts, they tend to use their network of friends and relatives. In other words, a network of people knowing each other creates the underlying infrastructure for any collective "online" information processing.

In the present paper we study the topology and dynamics of networks which are defined by interactions between social actors. Generally, we can model such a network with a graph where nodes represent people or groups of people and edges represent interactions and mutual knowledge between the actors. During the last decade, the theory of small worlds [4] and scale-free networks [5] have brought much attention to the general study of large networks which encounter dynamic changes in structure over time. Apart from applications in social sciences, the knowledge of structural features of large networks brought by these theories has many applications in information sciences, computer science, brain research, molecular biology and other fields.

To model a large group of interacting actors, one does not only need to express the fact that there is an interaction, but also in which direction the interaction occurs. Therefore it is necessary to introduce directional networks, which can be modeled by directed graphs. Recently Ahnert and Fink have experimentally studied directed networks [1] and introduced a quantity called clustering signature. Their experiments, on various sorts of large real world networks, have shown that the clustering signature is an important parameter which can structurally distinguish between different types of real networks. However, a complete analysis of the dynamics that lead to certain clustering signatures remains open.

As the name suggests, the clustering signature in a directed network describes how strong the nodes are clustered. It can be measured by counting occurrences of different types of directed triangles (see Figure 4), since their distributions characterize more complex clusters in the network. To understand the clustering signature, a hierarchy of simple objects and their distributions must be understood. To achieve this goal we study a hierarchy of relations among simple object distributions starting with nodes, arcs and wedges (see Figure 2).

Apart from statically understanding the topological building blocks of directed social networks, the dynamic development of the topology must be understood. Based on our understanding of the clustering signature we propose a prototypical stochastic process which models creation and movement of network links and allows to a large extent a control of the clustering signature. The design of this process is largely based on considerations about how social actors can interact to build the structure of the underlying network through which further social information processing occurs. Simulation results indicate that the building blocks of the process have a strong influence on the clustering signature. However, the exact understanding of this influence would need more theoretical research into process guarding equations and their solutions.

Our research concentrates on situations where the network is saturated in the number of interactions and participants. We think that it is very important to study the dynamics that occurs in equilibrium networks [6] where the number of nodes and edges does not grow significantly. The reason is that every network saturates after an initial explosive growth, however, later there is a restructuring process which largely defines the new structure of the network after a certain time. It is important to understand that the dynamics during the growth can deviate significantly from the one during the stable phase. The study of this class of processes was neglected in the first stage of modern network research, but nowadays, as networks which are already saturated are moving into the center of interest, the saturation effects are gaining importance [7].

In this paper we define a new equilibrium network process that has parameters which control the clustering character-

978-0-7695-3823-5/09 $26.00 © 2009 IEEE
DOI 10.1109/CSE.2009.273

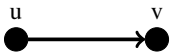

Fig. 1: An arc with head v and tail u.

istics in directed networks. First, we study the hierarchy of distributions which influence the clustering signature. Then we use this knowledge to extend an established process by a new building block which is motivated by actions that can be observed in social networks. The new building block allows to control the clustering characteristics in a wide range.

Our paper is organized as follows. In Section II we study the distribution hierarchy up to the triangles occurring in the clustering signature. In Section III, based on interactions between social actors we define a process which controls the clustering signature. In Section IV we simulate the process using various parameters to show that the clustering signature can be largely controlled by the parameters. Finally, we conclude suggesting the next steps which can contribute to further understanding of dynamic properties of the clustering signature.

II. Distribution Hierarchy

In this section we summarize the distribution hierarchy (for detailed definitions and proofs see [10]). We use the following basic notation. A directed network is defined by a set of nodes and a set of arcs where an arc (u, v) is a directed edge between a head node v and a tail node u. The indegree of a node v denoted by $deg^-(v)$ is the number of incoming arcs where as the outdegree $deg^+(v)$ is the number of outgoing arcs. We define the degree distribution $\overrightarrow{P}(k^+, k^-)$ as the probability that a random node has indegree k^- and outdegree k^+. The indegree distribution $\overrightarrow{P}^-(k)$ is completely defined by the degree distribution and can be computed by summing over all possible outdegrees.

To understand the network dynamics one has to study the probability that certain subgraphs occur in the network. Consider a small network object H, for such an object we can introduce a distribution that defines the probability that if we select an instance of H at random (from all subgraphs of the network which have structure H) it has a specific in- and outdegree for every node. As an example we look at the case where H is an arc (Figure 1). Then we define the arc distribution $\overrightarrow{P}_a(k^+, k^-)$ which denotes the probability that for a random arc, the head v has indegree k^- and the tail u has outdegree k^+. These distributions are sometimes related in the sense that it might be possible to compute one of them from one or more others. In this sense the distributions build a hierarchy and it is of high importance for the study of the clustering signature to understand how it is related to other lower order distributions.

A. Arc, Wedge and Triangle Distributions

Apart from the already defined degree, indegree, outdegree and arc distributions we will look at wedge and triangle distributions. Wedges are paths of length two which appear as three

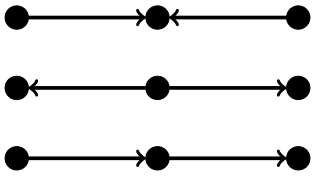

Fig. 2: The three different kinds of directed wedges.

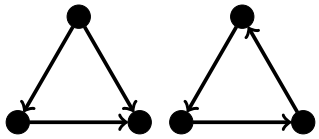

Fig. 3: The feedforward and the feedback loop.

different subgraphs (see Figure 2) in directed networks. The path (p-wedge) is a directed path of length two. The broadcast (b-wedge) is a middle node that has two outgoing arcs and the sink (s-wedge) is a middle node with two incoming arcs. There are two forms of directed triangles (Figure 3) to which we refer as the feedback loop (a directed cycle of length three) and the feedforward loop (a b-wedge where both end nodes are connected).

B. Clustering signature

The clustering coefficient in an undirected network is a quantity that describes the availability of triangles. It is well known that in social networks it is much larger then in other complex networks [8]. In directed networks there are two different types of neighbors (there is an in-neighborhood and a out-neighborhood) and they can be connected by two different arcs. Therefore we get the four different types of local clustering structures drawn in Figure 4.

In [1] Ahnert and Fink developed a generalization of the clustering coefficient in directed networks and called it the clustering signature.

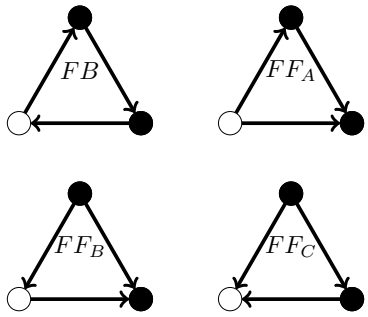

Fig. 4: The four directed triangles from the point of view of the white node. The clustering signature is a quantity which depends on their occurences in the network.

Definition II.1 (Clustering signature[1])**.** *The local clustering signature for vertex v is the four dimensional vector:*

$$C^{(i)} = (\frac{N_{FB}^{(i)}}{M_B^{(i)}}, \frac{N_{FF_A}^{(i)}}{M_A^{(i)}}, \frac{N_{FF_B}^{(i)}}{M_B^{(i)}}, \frac{N_{FF_C}^{(i)}}{M_C^{(i)}})$$

where $N^{(i)}$ is the number of triangles of a certain type in which the vertex participates and

$$M_B^{(i)} := \sum_{\substack{u \in \Gamma^+(i) \\ v \in \Gamma(i)}} (1 - \delta_{u,v}),$$

$$M_A^i := deg^+(i) \cdot (deg^+(i) - 1)$$

and

$$M_C^{(i)} := deg^-(i) \cdot (deg^-(i) - 1).$$

Here δ denotes the Kronecker delta and $deg^-(v)$ ($deg^+(v)$) the indegree (outdegree) of v . The global clustering signature is the average local clustering signature

$$C = \frac{1}{N} \sum_{i=1}^{N} C^{(i)}$$

and we also use the definition of the normalized clustering signature

$$\tilde{C} = \frac{1}{N} \sum_{i=1}^{N} \frac{(\frac{N_{FB}^{(i)}}{M_B^{(i)}}, \frac{N_{FF_A}^{(i)}}{M_A^{(i)}}, \frac{N_{FF_B}^{(i)}}{M_B^{(i)}}, \frac{N_{FF_C}^{(i)}}{M_C^{(i)}})}{(\frac{N_{FB}^{(i)}}{M_B^{(i)}} + \frac{N_{FF_A}^{(i)}}{M_A^{(i)}} + \frac{N_{FF_B}^{(i)}}{M_B^{(i)}} + \frac{N_{FF_C}^{(i)}}{M_C^{(i)}})}$$

.

In [9] Dorogovtsev defined clustering characteristics of undirected networks and derived their relations in correlated and uncorrelated networks. We will generalize one of this definitions to directed networks and show that it is equivalent to the clustering signature.

Definition II.2 (The in-outdegree dependent local clustering distribution)**.**

$$C(k^+, k^-) = \mathbb{E}[C^{(i)} | deg^+(i) = k^+, deg^-(i) = k^-]$$

$C(k^+, k^-)$ is a vector whose components are the expected relative number of closed loops of a certain type between the in-neighbors and the out-neighbors of a vertex with in-degree k^+ and out-degree k^-.
We can calculate the mean value over all $C(k^+, k^-)$:

$$\overline{C} = \sum_{k^+,k^-} \overrightarrow{P}_v(k^+, k^-) C(k^+, k^-) \tag{1}$$

In Observation II.1 we show that the two quantities defined in Definition II.1 and Definition II.2 are equivalent which means that the clustering signature is the natural generalization of the clustering coefficient to directed networks.

Observation II.1.

$$\overline{C} = C$$

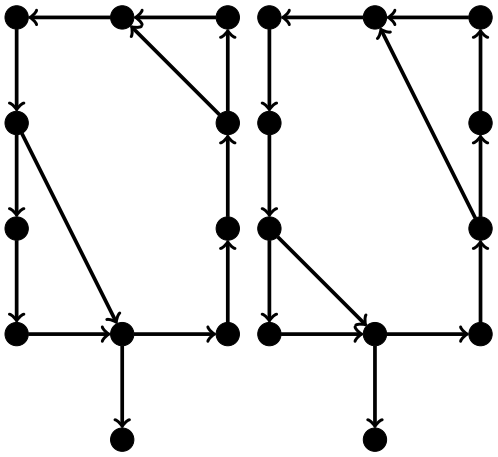

Fig. 5: Two networks with equal degree, indegree, outdegree, arc and wedge distributions. The clustering signatures are also equal for both networks, but the triangle distribution are different.

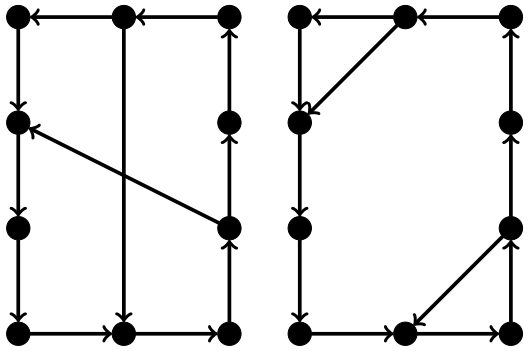

Fig. 6: Two networks with equal degree, indegree, outdegree, arc and wedge distribution but different clustering signature.

Proof:

$$\overline{C} = \sum_{k^+,k^-} \overrightarrow{P}_v(k^+, k^-) C(k^+, k^-)$$

$$= \sum_{k^+,k^-} \overrightarrow{P}_v(k^+, k^-) \cdot \mathbb{E}\left[C^{(i)} | deg^+(i) = k^+, deg^-(i) = k^-\right]$$

$$= \mathbb{E}\left[C^{(i)}\right]$$

$$= \frac{1}{N} \sum_{v \in V} \left(\frac{N_{FB}(v)}{M_B^{(v)}}, \frac{N_{FF_A}(v)}{M_A^{(v)}}, \frac{N_{FF_B}(v)}{M_B^{(v)}}, \frac{N_{FF_C}(v)}{M_C^{(v)}} \right) = C$$

■

C. The big picture

The complete relation hierarchy among all distributions up to triangles, which we formally studied in [10], is illustrated in Figure 7. From this picture we observe that the relations

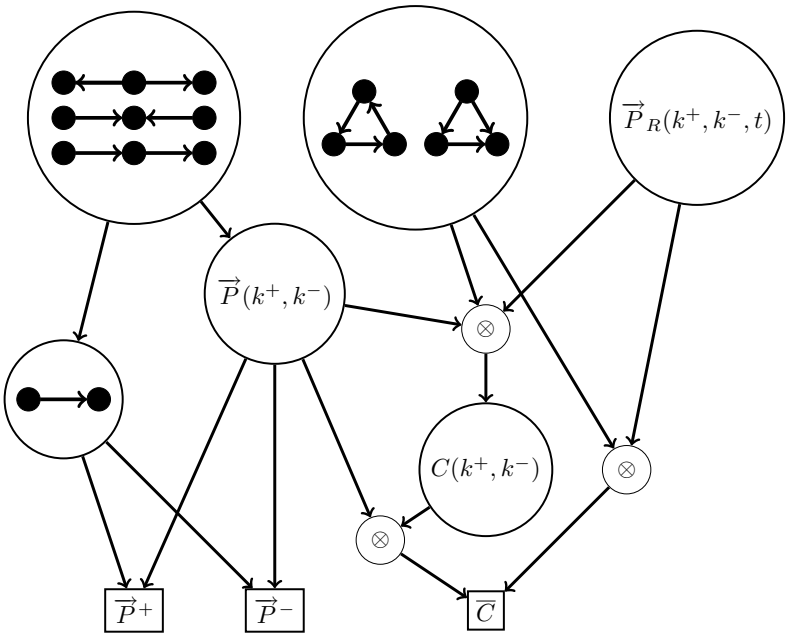

Fig. 7: The complete distribution hierarchy. An arrow from A to B means that B is uniquely defined by A. An arrow leading from $\otimes$ to B means that B can be calculated if we know all distributions which have an arrow to $\otimes$.

between indegree, outdegree, arc and wedge distribution (the leftmost part of the hierarchy in Figure 7) are almost like in undirected networks [11]. But to derive the degree distribution, we need at least the wedge distribution. The middle and the right part of the hierarchy does not exist in undirected networks at all, which shows additional order of complexity introduced by edge orientation. The general technique to prove that distribution P_A can be calculated from a set of distributions B, is to give a formula for P_A that depends only on distributions in B. Furthermore one proofs that P_A can not be derived from B by constructing two graphs for which all distributions in B are equivalent but the distribution P_A differs.

In the present paper we illustrate only the relevant relations, which provide a better understanding of the clustering signature. The counter example in Figure 6 shows that it is not possible to calculate the clustering signature from the degree, indegree, outdegree, arc and wedge distribution. This is not surprising because we cannot get any information about triangles from these distributions. A similar counter example (Figure 5) shows that the triangle distribution cannot be derived from all lower order distributions and the clustering signature. If we know the triangle distribution and a measure for the reciprocity ($\overrightarrow{P}_R(k^+,k^-,t)$) in the network, then we can calculate the global clustering signature [10]. If in addition the degree distribution is known, we are able to derive the local clustering signature and from this the global clustering signature.

III. A Social Network Process with Variable Clustering Signature

To understand the various phenomena in complex networks it has been useful to study stochastic processes which model the dynamics of the network and simulate the evolution of the phenomena. The two natural approaches to model changes in an equilibrium network are called edge rewiring [4] and vertex based addition and deletion of edges [11]. The first one is to choose an edge and one of its end nodes and to rewire the edge to a new node. Often a preference function that depends on the degree of the vertex is used to choose the second node. A well known process that uses this model is studied in [6]. The second approach operates in two different steps which are repeated with a certain probability each. The first step (removal) is to choose two nodes and remove the edge between those two nodes if they are connected. In the second step (addition) one chooses again two nodes and connects them if they are not already neighbors. By repeating both steps with the appropriate probability, the number of edges remains approximately constant.

The vertex based approach has several advantages. First, it always generates simple graphs while the edge based rewiring could generate multiple edges and self-loops (unless we use some constraints to avoid this events). But this is not the only advantage of the vertex based approach. In [11] it was shown that processes which are based on edge rewiring are very hard to analyze. It seems also more natural for the modeling of

various real world networks to add and remove new edges then to rewire them. For example a Twitter user might decide to follow a new blog or to unsubscribe to one of the blogs he is currently following, but these decisions are often not correlated.

The model we use here is based on VADE (vertex based addition and deletion of edges) proposed in [11]. Each iteration of the VADE consists of an addition step which is executed with probability p or a removal step executed with probability $(1-p)$. For the addition step, two vertices are chosen independently at random with probability proportional to some preference functions. If the chosen vertices are distinct and not connected, an edge between them is added. In the removal step two vertices are chosen with probability proportional to some other preference function and the edge between them is removed if it exists. Simulations show that the degree sequence of networks generated by VADE is scale free distributed for appropriate preference functions. But unfortunately there is no parameter that has a firsthand effect on the clustering coefficient. In most complex network this is not a problem as its clustering coefficient is understood as a side effect of the scale free degree distribution. For social networks however, the clustering is often much higher then in a random scale free graph [8].

In this work we define a process based on VADE that creates directed networks with a tunable clustering signature inspired by interactions between social agents in the Internet. Thereby we use nodes to model social agents and arcs to model the interactions between them. The generalization of VADE to directed networks is straightforward. The only thing we have to change is that the preference functions should depend on the in- and outdegree of the vertices. We extend VADE by introducing a new (third) step which models the dynamics that lead to the clustering structures observed in social networks [12].

We often use Twitter to illustrate how the relations between the social agents develop. Twitter is a very popular microblogging platform with more than 800000 users. The users can follow other users messages and this can be modeled with incoming edges. A user's actions are broadcasted to all users which follow him. There are not only persons present in Twitter but also institutions creating a heterogeneous network with various sorts of social agents.

In social networks there are basically two ways how the edges emerge. The first is that two vertices get connected independently of their neighbors. An example for this action in a Twitter network would be the circumstance in which user A googles for 'complex networks', finds a reference to a Twitter account and decides to follow this account. We will refer to this as the long jump because it connects two vertices from different areas of the network. So far all the processes we are aware of are based on long jumps [4], [11]. We claim that in most real world network a different effect called short jump exists. During a short jump two nodes which are endpoints of a common wedge establish a connection (note that they close a triangle in doing so). In the Twitter example user A can make a short jump by browsing the profile of an other user B whom he is following and deciding that he is interested in following user C whom user B is already following. In the social network of personal contacts a short jump is the introduction of two persons by a common acquaintance. We model the short jump as a substep of the addition step by choosing a node (with probability proportional to some preference function) and a random wedge through this node (where we have a parameter that defines with which probability we choose each kind of wedge). Then we connect the endpoints of this random wedge if they are not already connected. For the selection of the random wedge we use three different cases to which we assign a probability each. The first case is the broadcast wedge (selected with probability p_B) and the second the sink wedge (p_S). There are two possibilities to close a path wedge to a directed triangle and therefore we distinguish between closing it to a feedback loop (p_{PR}) or to a feedforward loop (p_P). The complete process is formally described in Process III.1.

Process III.1. *The following steps are repeated on a graph G in each discrete time unit. (Note that $p_P+p_{PR}+p_B+p_S=1$)*

1. *With probability p_{add} do the following:*
- ***a)*** *With probability $p_{add.sh.jmp}$:*
  - ***i)*** *Choose a vertex V_{middle} uniformly at random.*
  - ***ii)*** *With probability p_P (p_{PR}, p_B, p_S) choose a wedge $w=(V_i, V_{middle}, V_j)$ of type P (PR, B, S) at random.*
    If $V_i = V_j$ or there is an arc from V_i to V_j, skip the next step.
    Add an arc from V_i to V_j.
- ***b)*** *With probability $1-p_{add.sh.jmp}$*
  - ***i)*** *Choose a vertex V_i with a probability proportional to $f_{ao}(k)$.*
  - ***ii)*** *Choose a vertex V_j with a probability proportional to $f_{ai}(k)$.*
  - ***iii)*** *If $V_i = V_j$ or there is an arc from V_i to V_j, skip the next step.*
  - ***iv)*** *Add an arc from V_i to V_j.*

2. *With probability $1-p_{add}$ do the following:*
- ***i)*** *Choose a vertex V_i with a probability proportional to $f_{ro}(k)$.*
- ***ii)*** *Choose a vertex V_j with a probability proportional to $f_{ri}(k)$*
- ***iii)*** *If there is no arc from V_i to V_j, skip the next step.*
- ***iv)*** *Remove the arc from V_i to V_j.*

A. Arc stability

In what follows we will work out an approximation for p_{add} such that the number of arcs does not increase nor decrease over time, which means that the simulated network will not leave the equilibrium state. We assume that the probability for an arc (u,v) to be part of the network is equal for all vertices u, v. Thus, we have

$$p_r := \Pr[(u,v) \text{ can be removed}] = \frac{L}{N \cdot (N-1)}$$

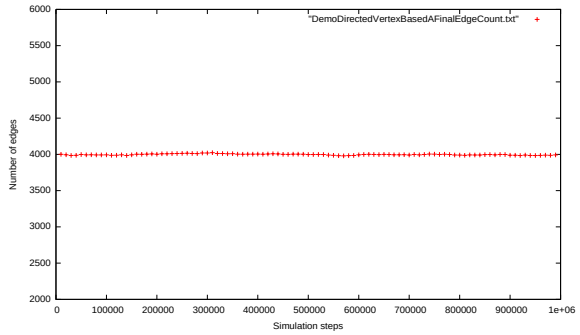

Fig. 8: The figure shows the simulation results of the process (for $p_{add.sh.jmp} = 0.1$), where the x axis represents the number of simulation steps and the y axis the number of arcs in the network.

and

$$p_a := \Pr[(u, v) \text{ can be added}] = 1 - p_r \ .$$

The goal is that the number of arcs becomes stable. Hence the probability of adding an arc should be equal to the probability of removing an arc. Therefore we require

$$pr \cdot (1 - p_{add}) = p_a \cdot p_{add} \ ,$$

which implies

$$p_{add} = p_r = \frac{L}{N \cdot (N-1)} \ . \tag{2}$$

The simulation results in Figure 8 suggest that this approximation leads to a stable amount of arcs in the network.

IV. Simulations

By looking at the simulation results one can get a deeper understanding of how the parameters of the process and the clustering signature of the created network are related. We describe the framework we used for the simulations first and discuss the results afterwards.

A. Simulation framework

For the following simulations we started from a random network with 1000 nodes and 4000 arcs. Therefore our approximation (Eq. 2) yields $p_{add} = 0.00404$. For the first simulations we calculated 10^6 steps per run and plotted the average over 5 runs. As preference functions we used

$$f_{ao}(k^+, k^-) = \begin{cases} 1000 & \text{if } k^+ \leq 1 \\ (k^+ + k^-) & \text{else} \end{cases} \ ,$$

$$f_{ai}(k^+, k^-) = \begin{cases} 1000 & \text{if } k^- \leq 1 \\ (k^+ + k^-) & \text{else} \end{cases}$$

and

$$f_{ro}(k^+, k^-) = f_{ri}(k^+, k^-) = 1.$$

Observe that for $p_{add.shrt.jmp} = 0$ the process is equivalent to VADE. If we define $p_{add.shrt.jmp} = 0.1$ then the distribution looks very similar (Figure 9) no matter what the wedge probabilities are. Please note that because of the boundary conditions $f_{ao}(0) = f_{ai}(0) = 1000$ there are almost surely no isolated vertices.

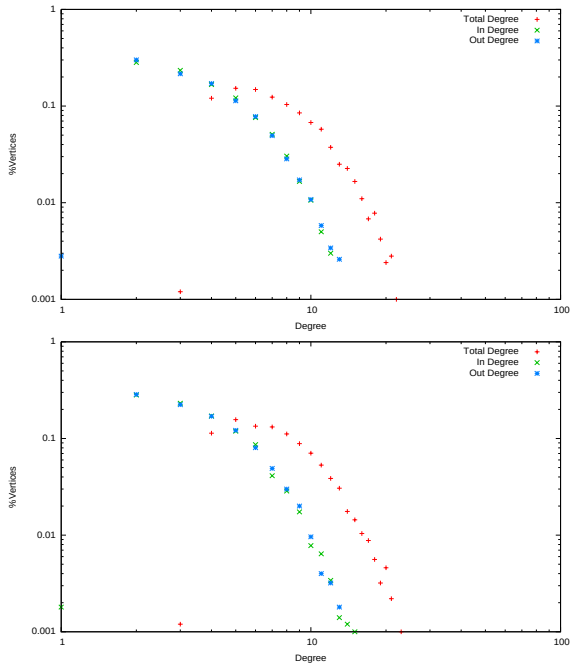

Fig. 9: This figure demonstrates that if we introduce a small jump probability of 0.1 then the degree distribution looks very much the same. The top picture shows the degree distribution after 1000000 steps with $p_{add.shrt.jmp} = 0$ and the bottom picture with $p_{add.shrt.jmp} = 0.1$. All axes are logarithmically scaled.

B. Discussion

The probability distribution that determines what kind of wedge we use in a short jump step is the parameter that is used to control the clustering signature. As soon as the probability that we select a directed path which is then closed to a feedback loop is larger then $1/4$, the FF component becomes dominating in the clustering signature. This is because each closed feedback loop improves the FF component of all three vertices contained in the triangle. But if we close a feedforward loop then for each node in the triangle, another component of the clustering signature increases. On the other hand, the simulations confirm that if we choose $p_p = p_{pr} = p_b = p_s = 0.25$ then all four components of the normalized clustering signature concentrate around the same value.

It is more interesting to study the boundary cases of the distribution where the probability for one wedge is set to 1 and for the others to 0. For B-wedges we observe that the FF_A component becomes dominating (Figure 10) because we improve in every short jump step the FF_A value of the selected node v_{middle}. By the same consideration we expect for $p_p = 1$ the FF_B and for $p_s = 1$ the FF_C to become dominant. For

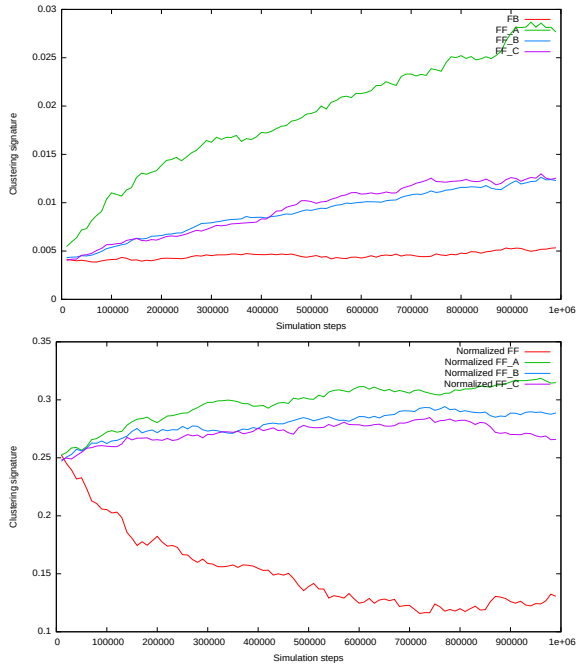

Fig. 10: Results of a simulation of the process with parameters $p_{add.sh.jmp} = 0.1$ and $p_b = 1$. In the top picture the absolute values of the clustering signature and in the bottom picture the normalized clustering signature is plotted against the number of steps.

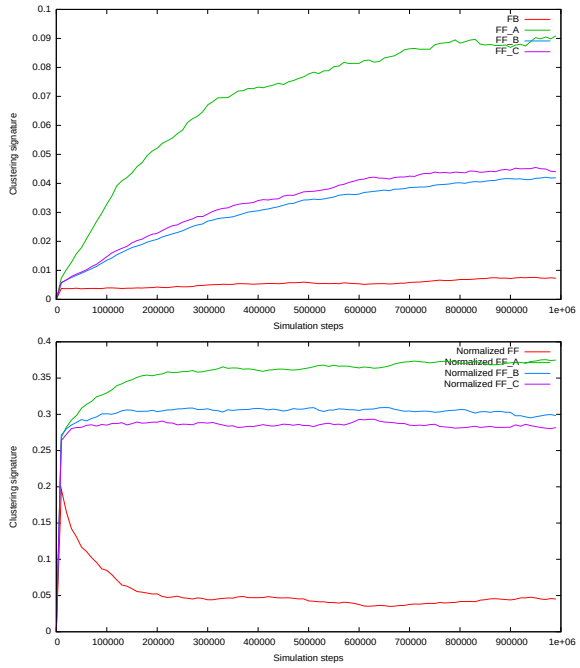

Fig. 11: Results of a simulation of the process with parameters $p_{add.sh.jmp} = 0.5$ and $p_b = 1$. In the top picture the absolute values of the clustering signature and in the bottom picture the normalized clustering signature is plotted against the number of steps.

$p_{pr} = 1$ of course the FB component grows (Figure 12) but the effect is much stronger as in the other case because all three vertices improve in their FB signature.

We are able to control the clustering signature in a wide range but in our simulations scenario, where the simulated networks have medium to low density and $p_{add.sh.jmp} = 0.1$, this effect is weaker for the normalized clustering signature (except for the feedforward loop). Let us examine this problem on the example in which we use only B-wedges to increase the FF_A component (Figure 10). If we close a B-wedge (v_l, v_{middle}, v_r) by adding the arc (v_l, v_r) then we increase FF_A at v_{middle} but most likely we also increase the FF_B signature at v_l and the FF_C signature at v_r. Those increase only little but with high probability the signature of a vertex is zero for all components (because the network is sparse). Now if we increase one component only slightly, this component becomes 1 in the normalized signature. On the other hand it seems reasonable to assume that the normalized clustering signature curve will show the same trends as the curve of the clustering signature if the network is dense enough such that for each vertex all components of the clustering signature are nonzero with high probability. We conclude that the density of the network has a non neglectable influence on the normalized clustering signature.

Another parameter which has a strong influence on the difference between the clustering signature and the normalized clustering signature is $p_{add.sh.jmp}$. This parameter is used to balance between long jump steps which are used to get the scale free degree distribution and short jump steps which increase the local clustering signatures. In a highly clustered network, the local clustering signature for an arbitrary node has almost surely only nonzero components. But this means, that some small increases in the feedforward components do not have a strong impact on the normalized clustering signature. For that reason $p_{add.sh.jmp}$ does not only affect the value of the local clustering signature but also its influence on the normalized clustering signature (compare Figure 10 and Figure 11).

V. Conclusion

The process proposed in our paper shows how to design mechanisms that do not only create scale free degree distributions but also specified clustering signatures. First we derived that the clustering signature is defined by the triangle distribution and the degree distribution. Therefore we knew that our process needs to change these two distributions to have an impact on the clustering signature. Then we used this knowledge and designed a process which closes wedges to triangles. We finally presented our simulation results which showed that by defining the probability with which each wedge type is selected, we are able to control the clustering signature.

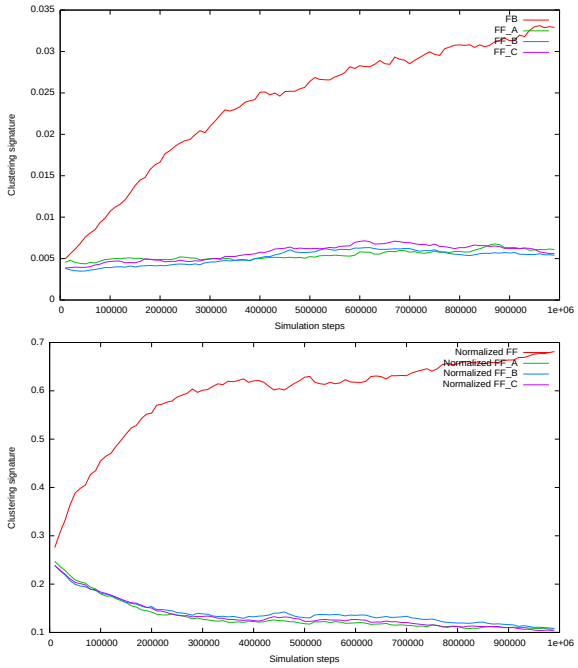

Fig. 12: Results of a simulation of the process with parameters $p_{add.sh.jmp} = 0.1$ and $p_{pr} = 1$. In the top picture the absolute values of the clustering signature and in the bottom picture the normalized clustering signature is plotted against the number of steps.

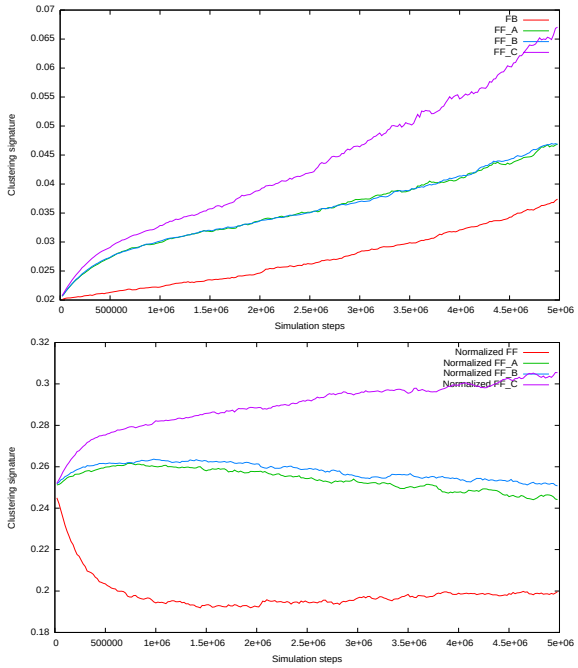

Fig. 13: Results of a simulation of the process with 20000 arcs and parameters $p_{add.sh.jmp} = 0.5$ and $p_s = 1$. In the top picture the absolute values of the clustering signature and in the bottom picture the normalized clustering signature is plotted against the number of steps.

An accurate comparison with real world networks of Twitter's size would need more research. The stochastic processes of the type we study have a very slow convergence. To simulate dense networks of sizes of order 10^4 - 10^5 new and probably parallel algorithms have to be developed. Figure 13 shows a simulation with 20000 arcs but it does not reach a stable state (the clustering signature is still increasing). Nevertheless, we can see that the curve of the normalized clustering signature approximates the pattern of the clustering signature better than in the earlier case with 4000 arcs (Figure 10).

We restricted our attention on simulations of the parameter boundary cases, where one of the wedge probabilities was 1 and the remaining three 0. A direction of further research would be to find out how this parameters could be combined to approximate a predefined clustering signature.

Another research question is to derive and solve the mean-field governing equation (master equation) of Process III.1. We think that the starting point is to find a closed form for the triangle distribution of VADE and from there to derive a closed form for the triangle distribution of Process III.1. Then for generalizing this results to directed networks one must use the relations derived in [10].

REFERENCES

[1] S. E. Ahnert and T. M. A. Fink, "Clustering signatures classify directed networks," *Phys. Rev.*, 2008.

[2] A. Z. Kamran, "Network analysis of global politics," *Proceedings of NetSci'08, International Workshop and Conference on Network Science and Its Applications, Norwich UK*, pp. 78–80, 2008.

[3] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaiten, "Computer networks as social networks: Collaborative work, telework, and virtual community," *Annu. Rev. Sociol.*, p. 213ñ238, 1996.

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–442, 1998.

[5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 509, 1999.

[6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks*. Oxford University Press, 2003.

[7] I. Farkas, M. Derenyi, G. Palla, and T. Vicsek, "Equilibrium statistical mechanics of network structures," *Lect. Notes Phys.*, vol. 650, 2004.

[8] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Physical Review E*, vol. 68, no. 036122, 2003.

[9] S. N. Dorogovtsev, "Clustering of correlated networks," *Physical Review E*, vol. 69, no. 027104, 2004.

[10] U. Peter and T. Hruz, "Distribution hierarchies in directed networks," *ETH Computer Science Department Technical Report*, no. 261, 2009.

[11] T. Hruz, M. Natora, and M. Agrawal, "Higher-order distributions and nongrowing complex networks without multiple connections," *Physical Review E*, vol. 77, no. 046101, 2008.

[12] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "Structure and evolution of blogspace," *Communications of the ACM*, 2004.