Published as Technical Report of the Department of Computer Science, 621, ETH Zurich, 2009

Distribution Hierarchies in Directed Networks

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

May 1, 2009

Abstract

Recently, Ahnert and Fink [AF08] showed that some classes of directed networks are cleanly separated in the space of the clustering signature. In this work we will study the relation hierarchy among subgraph distributions in directed networks and derive how the clustering signature fits into this hierarchy. Thereby we gather a fundamental understanding of the network dynamics and build a framework for the analysis of stochastic processes.

1 Introduction

Recently, there has been considerable interest in stochastic processes that generate small world networks [WS98] and networks with scale free degree distributions [BA99]. In [HNA08] the authors defined a process that creates equilibrium networks [FDPV04] without multiple edges and self-loops. Evans and Plato proved in [EP07] for an edge rewiring process that the degree sequence of the created network is scale free distributed. However, there are no formal proofs for the distribution of most other equilibrium network models known.

In [AF08] Ahnert and Fink analyzed 16 networks of five different types (social networks, genetic transcription networks, word adjacency networks, food webs and electric circuits). They showed that the clustering signature which generalizes the clustering coefficient for directed networks, cleanly separates these five classes. This means that there are structural differences between different types of directed networks.

In this report we define some structures which have influence on the clustering signature. Furthermore, we build a hierarchy by deriving the relations among all this structures and the clustering signature. In Section 1 we set up the necessary notation and terminology. In section 3 we add a few observations to the definitions. Section 4 establishes the hierarchy between the distributions in directed networks. In section 5 the relations between the distributions are derived while in the sixth section it is shown that in some special cases these relations are much simpler. The last section contains a short summary and an outlook on further research in this area.

2 Definitions, Distributions

A directed network is defined by a set of vertices V that are connected by arcs ($A \subseteq V \times V$). An arc (u, v) consists of a source vertex u and and a target vertex v. We denote $N := |V|$ as the number of vertices and $L := |A|$ as the number of arcs in the network. To analyze the clustering signature we are interested in the distributions of vertex degrees, arcs, directed wedges (two arcs that share a common vertex) and triangles.

2.1 Vertices

Definition 2.1 (In- and out-degree)**.** *To measure the number of arcs that have a vertex v as source vertex we define the out-degree*

$$deg^+(v) := |\{(v,u) \in A | u \in V\}|$$

and similar the in-degree

$$deg^-(v) := |\{(u,v) \in A | u \in V\}|.$$

Definition 2.2 (Degree probabilities)**.** *We denote $\overrightarrow{N}^+(k^+)$ as the number of vertices of out-degree k^+ and $\overrightarrow{N}^-(k^-)$ as the number of vertices of in-degree k^-. From this quantities we derive the probability that a vertex has out-degree k^+ or in-degree k^-:*

$$\overrightarrow{P}^-(k^-) := \frac{\overrightarrow{N}^-(k^-)}{N}$$

$$\overrightarrow{P}^+(k^+) := \frac{\overrightarrow{N}^+(k^+)}{N}$$

Sometimes we will also use the joint distribution of in-degree and out-degree. Therefore we define $\overrightarrow{N}_v(k^+, k^-)$ as the number of vertices of in-degree k^- and out-degree k^+. Hence

$$\overrightarrow{P}(k^+, k^-) := \frac{\overrightarrow{N}(k^+, k^-)}{N}.$$

Definition 2.3 (Mean degree)**.** *Now we define the mean in-degree, the mean out-degree and the mean total degree as*

$$\overline{k^+} := \frac{1}{N} \sum_{k^+} \overrightarrow{N}^+(k^+) \cdot k^+,$$

$$\overline{k^-} := \frac{1}{N} \sum_{k^-} \overrightarrow{N}^+(k^-) \cdot k^-,$$

and

$$\overline{k} := \frac{1}{N} \sum_{k^+,k^-} \overrightarrow{N}_v^+(k^+, k^-) \cdot (k^+ + k^-).$$

Note that whenever we write $\sum_{k^+}$, this means the summation of all positive values of k^+ which is equivalent to $\sum_{k^+=1}^{\infty}$.

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

2.2 Arcs

Definition 2.4 (Arc distribution)**.** *Let $\overrightarrow{L}(k^+, k^-)$ denote the number of arcs $a := (u, v)$ with $deg^+(u) = k^+$ and $deg^-(v) = k^-$. Then the arc distribution is defined by*

$$\overrightarrow{P}_a(k^+, k^-) := \frac{\overrightarrow{L}(k^+, k^-)}{L}$$

Sometimes we refer to k^- as the in-degree and to k^+ as the out-degree of arc a.

2.3 Reciprocity

Sometimes we will use the notion of reciproc arcs. An arc (u, v) is reciproc if and only if the arc (v, u) is also part of the network.

Definition 2.5 (Reciprocity distribution)**.** *The reciprocity distribution $\overrightarrow{P}_R(t|k^+, k^-)$ denotes the probability that a given vertex of in-degree k^- and out-degree k^+ has t reciproc arcs.*

2.4 Wedges

Wedges are paths of length two. In a directed network we have three different types of wedges. The directed path which is a path of length two, the broadcast wedge which consists of a central vertex and two outgoing arcs and the sink wedge that is defined by a middle vertex and two incoming arcs.

Definition 2.6 (P-wedge)**.** *We write $W_P(k_l^+, k_m^-, k_m^+, k_r^-)$ for the number of directed paths of length two with $deg^+(v_l) = k_l^+, deg^-(v_m) = k_m^-, deg^+(v_m) = k_m^+$ and $deg^-(v_r) = k_r^-$ where v_m is the middle vertex.*
Furthermore let W_P denote quantity of P-wedges in the network. Clearly the P-wedge distribution is defined by

$$P_P(k_l^+, k_m^-, k_m^+, k_r^-) := \frac{W_P(k_l^+, k_m^-, k_m^+, k_r^-)}{W_P}.$$

Definition 2.7 (B-wedge)**.** *Analogically let us denote $W_B(k_l^-, k_m^+, k_r^-)$ as the number of broadcast wedges with $deg^-(v_l) = k_l^-, deg^+(v_m) = k_m^+$ and $deg^-(v_r) = k_r^-$ where v_m is again the middle vertex.*
Then W_B denotes the quantity of B-wedges in the network and the B-wedge distribution is defined by

$$P_B(k_l^-, k_m^+, k_r^-) := \frac{W_B(k_l^-, k_m^+, k_r^-)}{W_B}.$$

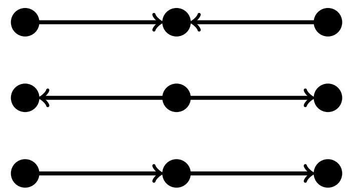

Figure 2: The three different kinds of directed wedges.

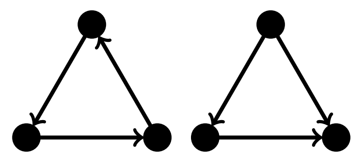

Figure 3: The feedback loop (left) and the feed-forward loop (right).

Definition 2.8 (S-wedge). *The number of sink wedges with* $deg^+(v_l) = k_l^+, deg^-(v_m) = k_m^-$ *and* $deg^+(v_r) = k_r^+$ *is* $W_C(k_l^+, k_m^-, k_r^+)$ *and the total number of S-wedges is* W_S*.*

$$P_S(k_l^+, k_m^-, k_r^+) := \frac{W_S(k_l^+, k_m^-, k_r^+)}{W_S}.$$

2.5 Directed Triangles

There are two types of directed triangles. The feedback loop (a directed loop) and the feed-forward loop.

Definition 2.9 (FeedBack loop FB). $T_{FB}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-)$ *denotes the number of feedback loops containing vertices of the given degrees and* T_{FB} *is the total quantity of feedback loops in the network.*
The feedback loop distribution is defined by

$$P_{FB}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-) := \frac{T_{FB}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-)}{T_{FB}}.$$

Definition 2.10 (FeedForward loop FF). $T_{FF}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-)$ *denotes the number of feedforward loops containing vertices of the given degrees and* T_{FF} *is the total quantity of feedforward loops in the network.*
The feedforward loop distribution is defined by:

$$P_{FF}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-) := \frac{T_{FF}(k_1^+, k_2^-, k_2^+, k_3^-, k_3^+, k_1^-)}{T_{FF}}.$$

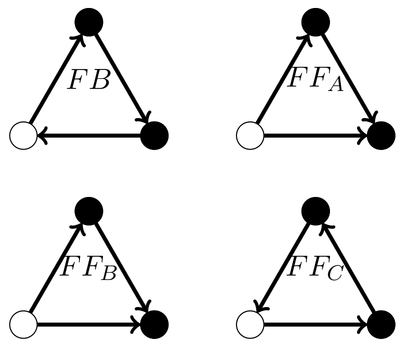

Figure 4: The four directed triangles from the point of view of the white node. The clustering signature is defined as a quantity which describes their occurence in the network.

2.6 Two ways of defining clustering signatures

The clustering signature is a quantity analogical to the clustering coefficient in undirected networks. The local clustering coefficient for a vertex v in an undirected network is defined as the number of connected pairs of neighbors of v divided by the number of pairs of neighbors of v. In directed networks there are two different types of neighbors and they can be connected by two different links. Therefore we get the four different types of local clustering structures drawn in Figure 4.

We will use the definition of the clustering signature from [AF08].

Definition 2.11 (Clustering signature[AF08]). *The local clustering signature for vertex v is the four dimensional vector:*

$$C^{(i)} = (\frac{N^{(i)}_{FB}}{M^{(i)}_{B}}, \frac{N^{(i)}_{FF_A}}{M^{(i)}_{A}}, \frac{N^{(i)}_{FF_B}}{M^{(i)}_{B}}, \frac{N^{(i)}_{FF_C}}{M^{(i)}_{C}})$$

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

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

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

and

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

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 [Dor04] Dorogovtsev defined certain clustering characteristics of undirected networks and derived the relations in correlated and uncorrelated networks. We will do the analog for directed networks.

Definition 2.12 (The in-outdegree dependent local clustering distribution)**.**

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

is a vector which has as components 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 $C(k^+, k^-)$:

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

3 Relations in directed networks

We use this section to explain the basic relations between the different distributions and also to calculate some probabilities which we need later.

3.1 Clustering signature

In section 2 we defined the mean clustering signature $\overline{C}$ and the global clustering signature C. Here we will show that this two definitions are equivalent.

Observation 3.1.

$$\overline{C} = C$$

Proof.

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

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

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

□

3.2 The average degree

The average in-degree is equal to the average out-degree because of $\sum_{v \in V} deg^+(v) = L = \sum_{v \in V} deg^-(v)$.

$$\overline{k^+} = \overline{k^-} = \frac{\overline{k}}{2} \tag{2}$$

Another useful equation is:

$$\overline{k^+} = \overline{k^-} = \left[\sum_{k^+, k'^-} \frac{1}{k^+} \overrightarrow{P}_A(k^+, k'^-) \right]^{-1} \tag{3}$$

Proof.

$$\overline{k^+} = \overline{k^-} = \frac{L}{N} = \frac{L}{\sum_{k^+} N^+(k^+)} = \frac{L}{\sum_{k^+} \frac{1}{k^+} \sum_{k'^-} L(k^+, k'^-)}$$
$$= \frac{L}{L \sum_{k^+, k'^-} \frac{1}{k^+} \overrightarrow{P}_A(k^+, k'^-)} = \left[\sum_{k^+, k'^-} \frac{1}{k^+} \overrightarrow{P}_A(k^+, k'^-) \right]^{-1}$$

□

4 Distribution hierarchy

In undirected networks there is a simple distribution hierarchy between vertex degree, edge and wedge where we can always derive the lower distribution from the higher one but not vice versa. In directed networks we have two types of vertex degree distributions and three types of wedge distributions. But is there still a simple hierarchy? And if there is one, what is its structure? To answer this question we will take a closer look at the relations between the different distributions.

4.1 Relation between vertex and arcs

The in- and the out-degree distributions ($\overrightarrow{P}^+(k^+)$ and $\overrightarrow{P}^-(k^-)$) are defined by their joint distribution $\overrightarrow{P}_v(k^+, k^-)$:

$$\overrightarrow{P}^+(k^+) = \sum_k \overrightarrow{P}_v(k^+, k)$$

and

$$\overrightarrow{P}^-(k^-) = \sum_k \overrightarrow{P}_v(k, k^-)$$

It is not possible to get the arc distribution from the in-out-degree distribution $\overrightarrow{P}_v(k^+, k^-)$ (and therefore also not from the in- or the out-degree distribution). A simple counter example is shown in Figure 5 where we have two networks with identical degree distributions but different arc distributions.

Like in the directed case, we can derive $\overrightarrow{P}^+(k^+)$ (and $\overrightarrow{P}^-(k^-)$) from $\overrightarrow{P_a}(k^+, k'^-)$.

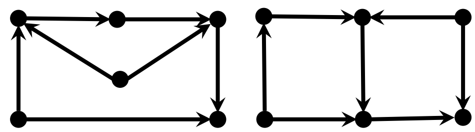

Figure 5: Two networks with identical degree distribution but different arc distribution.

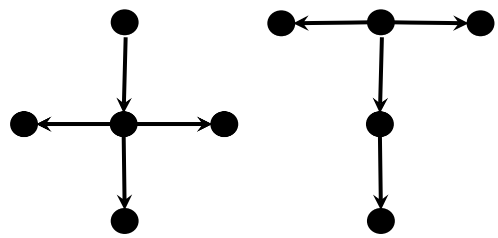

Figure 6: Two networks with identical arc distribution but different in-out-degree and wedge distribution.

Observation 4.1.

$$\overrightarrow{P}^{+}(k^{+}) = \frac{\overline{k^{+}} \sum_{k^{-}} \overrightarrow{P_a}(k^{+}, k^{-})}{k^{+}} \tag{4}$$

and

$$\overrightarrow{P}^{-}(k) = \frac{\overline{k^{-}} \sum_{k^{+}} \overrightarrow{P_a}(k^{+}, k^{-})}{k^{-}} \tag{5}$$

Proof.

$$\sum_{k^{-}} \overrightarrow{P_a}(k^{+}, k^{-}) = \frac{\text{number of arcs with outdegree } k^{+}}{L}$$

$$= \frac{\overrightarrow{N}^{+}(k^{+}) \cdot k^{+}}{\sum_{k^{+}} \overrightarrow{N}^{+}(k^{+}) \cdot k'^{+}} = \frac{\overrightarrow{P}^{+}(k^{+}) \cdot N \cdot k^{+}}{\sum_{k^{+}} \overrightarrow{P}^{+}(k^{+}) \cdot N \cdot k^{+}} = \frac{\overrightarrow{P}^{+}(k^{+}) \cdot k^{+}}{\overline{k^{+}}}$$

□

Another counter example (Figure 6) shows that it is in general not possible to derive the in-out degree distribution from the arc distribution.

4.2 Relation between arcs and wedges

4.2.1 Wedge distribution from arc distribution

Figure 6 shows two graphs which have the same arc distribution but different wedge distributions. This means that it is in general not possible to derive the wedge distribution from the arc distribution.

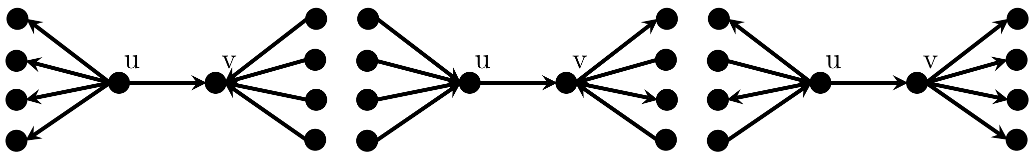

Figure 7: Counter examples

4.2.2 The weakly connected case

In a weakly connected network we are able to derive the arc distribution from the wedge distribution if and only if we know the distribution for all three types of wedge. This is what the following two lemmas state.

Lemma 4.1. *For each type of wedge there exist weakly connected networks in which it is impossible to calculate the arc distribution from the wedge distribution of this type.*

Proof. I show that for all three types of wedges there exist such an example (Figure 7 illustrates this counterexamples).

P-wedge: An arc (u, v) with $deg^-(u) = 0$ and $deg^+(v) = 0$ is not part of any P-wedge and can therefore not be counted by using the P-wedge distribution. Note that such arcs can occur in weakly connected networks.
B-wedges: An arc (u, v) with $deg^+(u) = 1$ is not part of any B-wedge and can therefore not be counted by using the B-wedge distribution.
S-wedges: An arc (u, v) with $deg^-(v) = 1$ is not part of any S-wedge and can therefore not be counted by using the S-wedge distribution.
This three examples are generic. We can construct a network with arbitrary many such arcs such that from a single wedge distributions it is not possible to derive the arc distribution. □

Lemma 4.2. *If we know the distributions for all three wedge types then the arc distribution is defined by*

$$\overrightarrow{P}_a(k^+, k'^-) = \frac{1}{L}\delta_{k^+,1} \cdot \delta_{k^-,1} W_P \left(\sum_{k',k''} \frac{\overrightarrow{P}_P(k^+, k^-, k', k'')}{k'} + \sum_{\substack{k' \neq k^-,\\ k'' \neq k^+}} \frac{\overrightarrow{P}_S(k', k'', k^+, k^-)}{k''} \right) \tag{6}$$

$$+ \frac{1}{L}\delta_{k^+,1} \cdot \left(1 - \delta_{k^-,1}\right) W_B \left(\sum_{k'} \frac{\overrightarrow{P}_B(k^-, k^+, k')}{k^+ - 1} \right)$$

$$+ \frac{1}{L}(1 - \delta_{k^+,1}) W_S \left(\sum_{k'} \frac{\overrightarrow{P}_S(k^+, k^-, k')}{k^- - 1} \right)$$

Proof. The following table shows for which in- and out-degree values of an arc we can use which kind of wedge distribution:

k^+	k^-	Wedge type	$\delta_{k^+,1} \cdot \delta_{k^-,1}$	$(1 - \delta_{k^+,1})$	$\delta_{k^+,1} \cdot (1 - \delta_{k^-,1})$
1	1	P	1	0	0
1	> 1	S	0	0	1
> 1	1	B	0	1	0
> 1	> 1	B	0	1	0

Counting arcs with in-degree 1 and out-degree 1:

$$\overrightarrow{L}(k^+, k^-) = \sum_{k',k''} \frac{W_P(k^+, k^-, k', k'')}{k''} + \sum_{k' \neq k^-, k'' \neq k^+} \frac{W_A(k', k'', k^+, k^-)}{k''}$$

Counting arcs with out-degree 1 and in-degree at least 2:

$$\overrightarrow{L}(k^+, k^-) = \sum_{k'} \frac{W_S(k^+, k^-, k')}{k^- - 1}$$

Counting arcs with out-degree at least 2:

$$\overrightarrow{L}(k^+, k^-) = \sum_{k'} \frac{W_B(k^-, k^+, k')}{k^+ - 1}$$

Hence we can add it all together

$$\begin{aligned}\overrightarrow{P}_a(k^+, k'^-) = \frac{1}{L}\delta_{k^+,1} \cdot \delta_{k^-,1} W_P \left(\sum_{k',k''} \frac{\overrightarrow{P}_P(k^+, k^-, k', k'')}{k'} + \sum_{\substack{k' \neq k^-,\\ k'' \neq k^+}} \frac{\overrightarrow{P}_S(k', k'', k^+, k^-)}{k''} \right) \\ + \frac{1}{L}\delta_{k^+,1} \cdot \left(1 - \delta_{k^-,1}\right) W_B \left(\sum_{k'} \frac{\overrightarrow{P}_B(k^-, k^+, k')}{k^+ - 1} \right) \\ + \frac{1}{L}(1 - \delta_{k^+,1}) W_S \left(\sum_{k'} \frac{\overrightarrow{P}_S(k^+, k^-, k')}{k^- - 1} \right)\end{aligned}$$

□

4.2.3 The strongly connected case

In a strongly connected network every vertex has at least one incoming and one outgoing arc. Hence there is an A-wedge trough every arc and therefore we can count the arcs by using only A-wedges:

$$\overrightarrow{L}(k^+, k^-) = \sum_{k',k''} \frac{W_A(k^+, k^-, k', k'')}{k'}$$

4.3 Relation between wedges and in-outdegree distribution

The wedge distribution gives us enough new information to calculate the in-outdegree distribution (remember that Figure 6 is a counter example that this is not possible from the arc degree distribution).

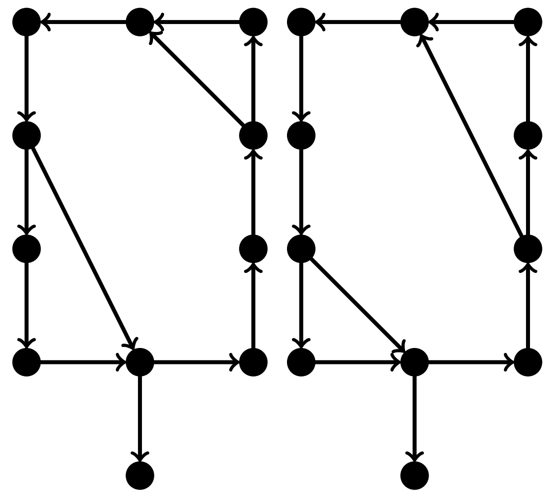

Figure 8: Two networks with equal degree, indegree, outdegree, arc and wedge distributions. The clustering signatures are also equal for both networks, but the triangle distribution are different.

Observation 4.2. *For $k^+ \neq 0$ and $k^- \neq 0$ we get the joint distribution of in and out degree by looking at all directed paths:*

$$\overrightarrow{N}_v(k^+, k^-) = \sum_{k,k'} \frac{W_p(k, k^+, k^-, k')}{k^+ \cdot k^-}$$

and the border can be derived from the arc distribution (which again can be calculated from the wedge distribution):

$$\overrightarrow{N}_v(k^+, 0) = \sum_{k} \frac{\overrightarrow{L}(k^+, k)}{k^+} - \sum_{k \neq 0} \overrightarrow{N}_v(k^+, k),$$

$$\overrightarrow{N}_v(0, k^-) = \sum_{k} \frac{\overrightarrow{L}(k, k^-)}{k^+} - \sum_{k \neq 0} \overrightarrow{N}_v(k, k^-)$$

4.4 Relation between triangles and clustering signature

Observation 4.3. *It is in general not possible to derive the triangle distribution from the clustering signature and all lower degree distributions (degree, arc, wedge).*

Proof. The two networks in Figure 8 have the same clustering signatures and the same lower degree distributions. But the triangle distribution is different because in the right example we have a triangle which contains a vertex of in-degree two and out-degree two. □

Observation 4.4. *From the triangle distribution, the in-out-degree distribution and the reciprocity distribution, we can get the local clustering distribution:*

$$C_{FB}(k^+, k^-) = \frac{\sum_{k_2^+, k_2^-, k_3^+, k_3^-} T_{FB}(k^+, k_2^-, k_2^+, k_3^-, k_3^+, k^-)}{\overrightarrow{N}_v(k^+, k^-) \sum_t \overrightarrow{P}_r(t|k^+, k^-)(k^+k^- - 1)},$$

$$C_{FF_A}(k^+,k^-) = \frac{\sum_{k_2^+,k_2^-,k_3^+,k_3^-} T_{FF}(k^+,k_2^-,k_2^+,k_3^-,k_3^+,k^-)}{\overrightarrow{N}_v(k^+,k^-)k^+(k^+-1)},$$

$$C_{FF_B}(k^+,k^-) = \frac{\sum_{k_1^+,k_1^-,k_3^+,k_3^-} T_{FF}(k_1^+,k^-,k^+,k_3^-,k_3^+,k_1^-)}{\overrightarrow{N}_v(k^+,k^-)\sum_t \overrightarrow{P}_r(t|k^+,k^-)(k^+k^- - 1)},$$

$$C_{FF_C}(k^+,k^-) = \frac{\sum_{k_1^+,k_1^-,k_2^+,k_2^-} T_{FF}(k_1^+,k_2^-,k_2^+,k^-,k^+,k_1^-)}{\overrightarrow{N}_v(k^+,k^-)k^-(k^- - 1)}.$$

Hence we can calculate the global clustering signature from the triangle distribution and the reciprocity distribution:

$$\overline{C} = \frac{1}{N} \sum_{\substack{k_1^+,k_1^-,k_2^+,\\ k_2^-,k_3^+,k_3^-}} \begin{pmatrix} \frac{T_{FB}(k_1^+,k_2^-,k_2^+,k_3^-,k_3^+,k_1^-)}{\sum_t \overrightarrow{P}_r(t|k_1^+,k_1^-)(k_1^+k_1^- - 1)} \\ \frac{T_{FF_A}(k_1^+,k_2^-,k_2^+,k_3^-,k_3^+,k_1^-)}{k_1^+(k_1^+ - 1)} \\ \frac{T_{FF_B}(k_1^+,k_2^-,k_2^+,k_3^-,k_3^+,k_1^-)}{\sum_t \overrightarrow{P}_r(t|k_1^+,k_1^-)(k_1^+k_1^- - 1)} \\ \frac{T_{FF_C}(k_1^+,k_2^-,k_2^+,k_3^-,k_3^+,k_1^-)}{k_1^-(k_1^- - 1)} \end{pmatrix}^T$$

4.5 Relation between triangles and lower order distributions

In Observation 4.3 we already saw that we can not compute the triangle distribution from the lower order distributions. On the contrary we can not derive a lower order distribution say the wedges from the triangle distribution. An intuitive argument is that from the triangle distribution we can only count objects that are contained in triangles. Therefore it is easy to find arbitrary large example networks which have the same triangle distribution but different lower order distributions.

4.6 Relation between clustering signature and lower oder distributions

Observation 4.5. *It is in general not possible to derive the clustering signature from all lower degree distributions (degree, arc, wedge).*

Proof. The two networks in Figure 9 have the same lower degree distributions. But one can verify that the clustering signature is different. □

4.7 The relation of the reciprocity distribution

Observation 4.6. *It is in general not possible to derive the reciprocity distribution from the degree, arc, wedge and triangle distributions and the clustering signature..*

Proof. The two networks in Figure 10 have the same degree, indegree, outdegree, arc, wedge and triangle distributions and an equivalent clustering signature. However, we observe that the reciprocity distribution is different. □

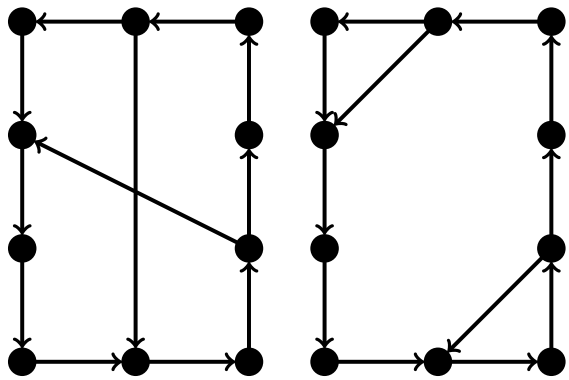

Figure 9: Two networks with equal degree, indegree, outdegree, arc and wedge distribution but different clustering signature.

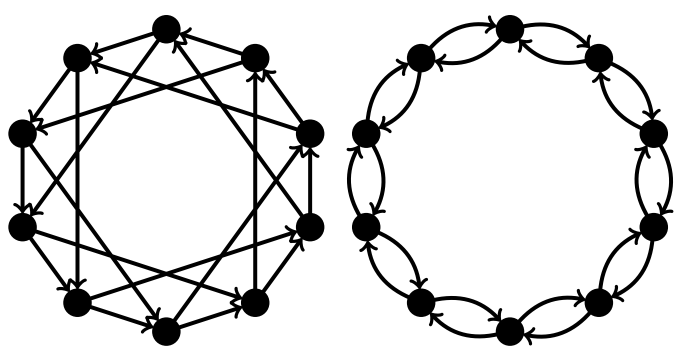

Figure 10: Two networks with equal degree, indegree, outdegree, arc, wedge and triangle distribution but different reciprocity distributions.

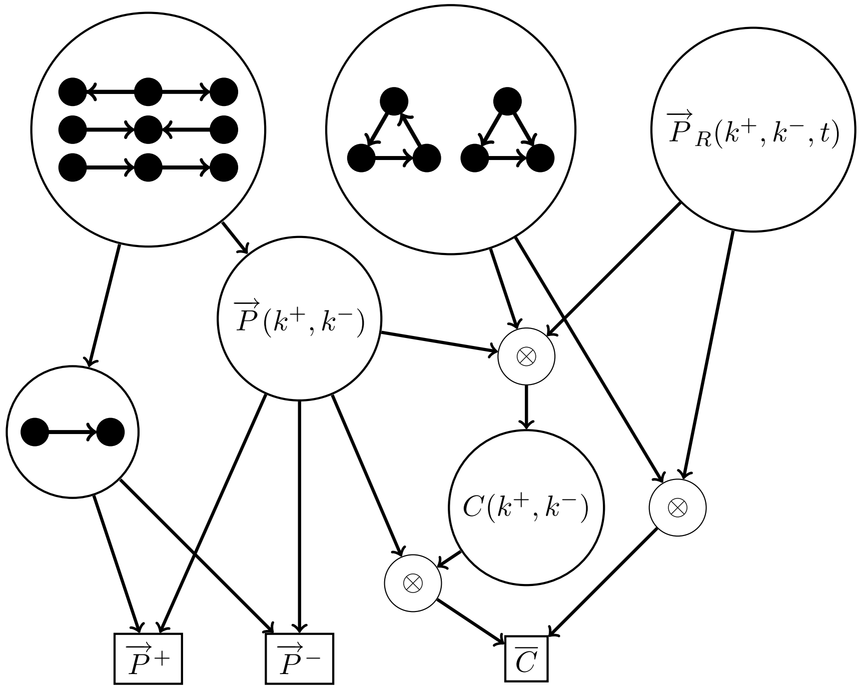

Figure 11: The complete distribution hierarchy. An arrow from A to B means that B is uniquely defined by A. An arrow leading from $\otimes$ to B means that B can be calculated if we know all distributions which have an arrow to $\otimes$.

4.8 The big picture

All the relations derived so fare are illustrated in Figure 11. This picture might help a lot for the analysis of a process. If we are interested in the clustering signature of the network then it suffices to find a closed formula for the triangle distribution (which might be easier to derive) and one for the reciprocity distribution. The main difference between directed and undirected networks is the joint distribution of in- and out-degree which is present only in directed networks. This joint distribution can only be derived from the wedge distribution. If we limit our attention on the lower degree distributions (no triangles and clustering signature) then the hierarchy looks very similar as in the undirected case. The arcs can be calculated from the wedges and the vertex in-/out-degrees from the arcs. Therefore it suffices to find a closed formula for the wedge distribution to analyze all this distributions.

5 Important probabilities

We use this section to collect the calculation of some probabilities which are later used. The reader may skip this chapter because we will always put a reference whenever we use one of this probabilities.

5.1 Selecting arcs at random

When it comes to the analysis of a process we are often interested in the probability that an uniformly at random selected arc (u, v) is of out-degree k:

Observation 5.1.

$$Pr[deg^+(u) = k^+] = \frac{k^+ \cdot \overrightarrow{P}^+(k^+) \cdot N}{L} = \frac{2 \cdot \overrightarrow{P}^+(k^+) \cdot k^+}{\overline{k}} = \frac{\overrightarrow{P}^+(k^+) \cdot k^+}{\overline{k^+}} \tag{7}$$

here we used $\overline{k} = \frac{2L}{N}$ *and* $\overline{k} = 2\overline{k^+}$*. With similar argumentation we can also get*

$$Pr[deg^-(v) = k^-] = \frac{\overrightarrow{P}^-(k^-) \cdot k^-}{\overline{k^-}} \tag{8}$$

5.2 Conditional arc degree probability

We define the probability that the in-degree of an arc with out-degree k^+ is k^- as

$$\overrightarrow{P}^{(-|+)}(k^-|k^+) = \frac{\overrightarrow{P}_a(k^+, k^-)}{\sum_k \overrightarrow{P}_a(k^+, k)} = \frac{\overline{k^+} \cdot \overrightarrow{P}_a(k^+, k^-)}{k^+ \cdot P^+(k^+)} \tag{9}$$

and the reverse as

$$\overrightarrow{P}^{(+|-)}(k^+|k^-) = \frac{\overrightarrow{P}_a(k^+, k^-)}{\sum_k \overrightarrow{P}_a(k, k^-)} = \frac{\overline{k^-} \cdot \overrightarrow{P}_a(k^+, k^-)}{k^- \cdot P^-(k^-)} \tag{10}$$

where we used Eq. 4 and 5 for the second equality.

6 Correlation

6.1 B-uncorrelated

In the classical random graph model the degree sequence is Poisson distributed and every edge (or arc in a directed network) has the same probability. This model does not represent the real world networks very well. In such networks we have certain types of correlations. But if we neglect some of this correlation then the calculation can become much easier. In undirected networks we are typically concerned about edge correlation and we call an undirected network uncorrelated if and only if $P(k, k') = \frac{k \cdot P(k) \cdot k' \cdot P(k')}{\overline{k}^2}$ holds for all k and k'. In directed networks there are more types of correlation which we have to take care of. Therefore we call a network B-uncorrelated if has no correlation of type B.

Definition 6.1 (in-out-degree-uncorrelated)**.** *We call a network inout-uncorrelated if the in- and the out-degree distribution are independent. Then we can calculate the joint distribution of in- and out-degree:*

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

.

Let us quickly repeat the concept of half edges and adapt it to directed networks. If we talk about half edges in undirected network then we see an edge $\{u, v\}$ as two half edges of which one is connected to vertex u and the other one to vertex v. In a directed network we can divide an arc (u, v) into a head that is connected to v and a tail that is connected to u.

Definition 6.2 (Arc-uncorrelated)**.** *Arc-uncorrelated means that the probability for a head at v and a tail at $u$to be part of the same arc is uniformly distributed over all heads and tails. Therefore we can compute the probability that an arc has out-degree k^+ and in-degree k^-*

$$\overrightarrow{P}_a(k^+, k^-) = \frac{k^+ \cdot P^+(k^+) \cdot k^- \cdot P^-(k^-)}{\overline{k^+} \cdot \overline{k^-}} \tag{11}$$

The natural step to expand this definition from the arc to wedges is to call a network wedge-uncorrelated if the event that two arcs of degree (k^+, k^-) and (k'^+, k'^-) are connected to a vertex v is independent of the degrees of the arcs.

Definition 6.3 (Wedge-uncorrelated)**.** *Lets define a network as P-wedge-uncorrelated if for all k_l^+, k_m^-, k_m^+ and k_r^-*

$$\begin{aligned} P_p(k_l^+, k_m^-, k_m^+, k_r^-) &= \overrightarrow{P}_v(k_m^-, k_m^+) \cdot \overrightarrow{P}^{(+|-)}(k_l^+|k_m^-) \cdot \overrightarrow{P}^{(-|+)}(k_r^-|k_m^+) \\ &= \frac{\overrightarrow{P}_v(k_m^+, k_m^-) \cdot \overrightarrow{P}_a(k_l^+, k_m^-) \cdot \overrightarrow{P}_a(k_m^+, r_m^-) \cdot \overline{k^+} \cdot \overline{k^-}}{k_l^+ \cdot k_r^- \cdot P^+(k_l^+) \cdot P^-(k_r^-)}. \end{aligned}$$

Furthermore a network is B-wedge-uncorrelated if for all k_l^-, k_m^+ and k_r^-

$$P_B(k_l^-, k_m^+, k_r^-) = \overrightarrow{P}_a(k_m^+, k_l^-) \cdot \overrightarrow{P}^{(-|+)}(k_r^-|k_m^+)$$

and S-wedge-uncorrelated if for all k_l^+, k_m^- and k_r^+

$$P_S(k_l^+, k_m^-, k_r^+) = \overrightarrow{P}_a(k_l^+, k_m^-) \cdot \overrightarrow{P}^{(+|-)}(k_r^+|k_m^-).$$

I will use the term wedge-uncorrelated for networks that are P-, B- and S-wedge uncorrelated.

In most of the real world networks the clustering is much larger than expected. This means that two vertices which share a common neighbor are more often connected then two arbitrary vertices. The probability that for a vertex u of out-degree k^+ and a vertex v of in-degree k^- the arc (u, v) is part of the network is

$$\left(\overrightarrow{P}^{(-|+)}(k^-|k^+) \cdot \frac{k^-}{N^-(k^-)k^-}\right) \cdot k^+.$$

This is the probability that a tail at u is connected to a head at v times the number of tails at u. We want do define triangle-uncorrelated in such a way that a network is triangle uncorrelated if and only if for every wedge the probability that there is an arc that closes the wedge to a triangle is the probability of this arc calculated above.

Definition 6.4. *A network is called triangle-uncorrelated if for a all wedges on the vertices u, v, w (where v is the middle vertex) the probability that an arc (u, w) exists is exactly*

$$\left(\overrightarrow{P}^{(-|+)}(deg^-(w)|deg^+(u)) \cdot \frac{deg^-(w)}{N^-(deg^-(w)) \cdot deg^-(w)}\right) \cdot deg^+(u).$$

6.2 Clustering in triangle-uncorrelated networks

In Chapter 3 we showed that it is in general not possible to derive the triangle distribution or clustering signature from the lower order distributions. But if we assume that there is no correlation on the level of triangles then we can derive the probability of triangles. We present the calculation only for the feedback loop but it is clear that one could derive the formulas for the other loops in the same way. The in-outdegree dependent local feedback clustering signature for a vertex v of in-degree k^+ and out-degree k^- is the probability that a fixed head at v and a fixed tail at v are part of a feedback loop. Here $\Gamma(v)$ denotes the neighborhood of vertex v.

$$C_{FF}(k^+,k^-) = \frac{\langle m_{FB}(k^+,k^-)\rangle}{k^+\cdot k^-} \tag{12}$$

$$= \Pr[(u,w) \text{ in the network}|\ u\in\Gamma^+(v), w\in\Gamma(v) \text{ fixed for } v \text{ of degree } (k^+,k^-)]$$

$$= \sum_{\substack{u^+,u^-\\w^+,w^-}} \overrightarrow{P}^{(-|+)}(u^-|k^+)\overrightarrow{P}^{(+|-)}(w^+|k^-)\overrightarrow{P}_v(u^+|u^-)\overrightarrow{P}_v(w^-|w^+)\times$$

$$\left(\overrightarrow{P}^{(-|+)}(w^-|u^+)\frac{1}{NP^-(w^-)}u^+\right)$$

To calculate the probability we have to fix the degrees of u and w and sum over all conditional probabilities for the different values of the degrees of u and w. $\overrightarrow{P}^{(-|+)}(u^-|k^+)$ and $\overrightarrow{P}^{(+|-)}(w^+|k^-)$ are the probabilities that u has in-degree u^- and w has out-degree w^+. Then $\overrightarrow{P}_v(u^+|u^-)$ and $\overrightarrow{P}_v(w^-|w^+)$ denote the probabilities that u has out-degree u^+ and w has in-degree w^-. The last part of the product is the probability that there is an arc from u to w. We can get this straight from the definition of triangle-uncorrelated.
We can insert Eq. (9) and (10) into (12) and get

$$C(k^+,k^-) = \frac{\overline{k^+}^3}{N\cdot k^+\cdot k^-\cdot P^+(k^+)\cdot P^-(k^-)} \tag{13}$$

$$\cdot \sum_{u^+,u^-,w^+,w^-} \frac{\overrightarrow{P}_v(u^+|u^-)\overrightarrow{P}_v(w^-|w^+)\overrightarrow{P}_a(k^+,u^-)\overrightarrow{P}_a(w^+,k^-)\overrightarrow{P}_a(u^+,w^-)}{P^-(w^-)}$$

this means that in an arc-uncorrelated network we get

$$C(k^+,k^-) = \frac{1}{N\cdot\overline{k^+}^3}\cdot\sum_{\substack{u^+,u^-\\w^+,w^-}} \overrightarrow{P}_v(u^+,u^-)\overrightarrow{P}_v(w^+,w^-)P^+(u^+)u^+u^-w^+w^-. \tag{14}$$

Here we can already see that the local clustering coefficient for the feedback loop does not depend on the in- and out-degree of the vertex when the network is arc-uncorrelated and triangle-uncorrelated.
For completeness we present here also the other clustering signatures in triangle-uncorrelated networks.

Lemma 6.1. *In a triangle-uncorrelated network we can calculate the local clustering signature from the degree distribution and arc distribution:*

$$C_{FB}(k^+, k^-) = \frac{\langle m_{FB}(k^+, k^-)\rangle}{k^+ \cdot k^-}$$

$$= \sum_{\substack{u^+,u^- \\ w^+,w^-}} \overrightarrow{P}^{(-|+)}(u^-|k^+)\overrightarrow{P}^{(+|-)}(w^+|k^-)\overrightarrow{P}_v(u^+|u^-)\overrightarrow{P}_v(w^-|w^+)\times$$

$$\left(\overrightarrow{P}^{(-|+)}(w^-|u^+)\frac{1}{NP^-(w^-)}u^+\right),$$

$$C_{FF_A}(k^+, k^-) = \frac{\langle m_{FF_A}(k^+, k^-)\rangle}{k^+ \cdot k^-}$$

$$= \sum_{\substack{u^+,u^- \\ w^-}} \overrightarrow{P}^{(-|+)}(u^-|k^+)\overrightarrow{P}^{(-|+)}(w^-|k^+)\overrightarrow{P}_v(u^+|u^-)\left(\overrightarrow{P}^{(-|+)}(w^-|u^+)\frac{1}{NP^-(w^-)}u^+\right),$$

$$C_{FF_B}(k^+, k^-) = \frac{\langle m_{FF_B}(k^+, k^-)\rangle}{k^+ \cdot k^-}$$

$$= \sum_{\substack{u^+ \\ w^-}} \overrightarrow{P}^{(+|-)}(u^+|k^-)\overrightarrow{P}^{(-|+)}(w^-|k^+)\left(\overrightarrow{P}^{(-|+)}(w^-|u^+)\frac{1}{NP^-(w^-)}u^+\right),$$

$$C_{FF_C}(k^+, k^-) = \frac{\langle m_{FF_C}(k^+, k^-)\rangle}{k^+ \cdot k^-}$$

$$= \sum_{\substack{u^+ \\ w^+,w^-}} \overrightarrow{P}^{(+|-)}(u^+|k^-)\overrightarrow{P}^{(+|-)}(w^+|k^-)\overrightarrow{P}_v(w^-|w^+)\times$$

$$\left(\overrightarrow{P}^{(-|+)}(w^-|u^+)\frac{1}{NP^-(w^-)}(u^+ - 1)\right).$$

7 Conclusion

The relations derived in this report can be helpful for the design and analysis of stocastic processes in directed networks. We derived the complete hierarchy among the distributions of vertex degrees, arcs, wedges and triangles. Furthermore, we introduced a clustering distribution which might often be more handy then the local clustering signature. We showed that the clustering distribution is equivalen to the clustering signature and derived its relation to all the previous distributions and also to the reciprocity distribution. At the end we defined some formal restrictions on the correlation between higher order distributions and pointed out that under the assumption of such restrictions some complex relations can be simplified.

As a further direction of research we suggest the design of an equilibrium process which operates on directed networks and has some parameters which allow a control of the clustering signature.

References

[AF08] S. E. Ahnert and T. M. A. Fink. Clustering signatures classify directed networks. *Phys. Rev.*, 2008.

[BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(509), 1999.

[Dor04] S. N. Dorogovtsev. Clustering of correlated networks. *Physical Review E*, 69(027104), 2004.

[EP07] T. S. Evans and A. D. K. Plato. Exact solution for the time evolution of network rewiring models. *Physical Review E*, 75(056101), 2007.

[FDPV04] I. Farkas, M. Derenyi, G. Palla, and T. Vicsek. Equilibrium statistical mechanics of network structures. *Lect. Notes Phys.*, 650, 2004.

[HNA08] T. Hruz, M. Natora, and M. Agrawal. Higher-order distributions and nongrowing complex networks without multiple connections. *Physical Review E*, 77(046101), 2008.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.