*Proceedings of EPNACS 2011 within ECCS'11*
*Emergent Properties in Natural and Artificial Complex Systems*
*Vienna, Austria - September 15, 2011*

# EFFECTS OF TIME-DEPENDENT EDGE DYNAMICS ON PROPERTIES OF CUMULATIVE NETWORKS

Richard O. Legendi and Laszlo Gulyas *†‡

**Abstract.** Inspecting the dynamics of networks opens a new dimension in understanding the interactions among the components of complex systems. Our goal is to understand the baseline properties to be expected from elementary random changes over time, in order to be able to assess the effects found in longitudinal data.

In our earlier work, we created elementary dynamic models from classic random and preferential networks. Focusing on edge dynamics, we defined several processes changing networks of fixed size. We applied simple rules, including random, preferential or assortative modification of existing edges - or a combination of these. Starting from initial Erdos-Renyi or Barabasi-Albert networks, we examined various basic network properties (e.g., density, clustering, average path length, number of components, degree distribution) of both snapshot and cumulative networks (of various lengths of aggregation time windows). In the current paper, we extend this line of research by applying time-dependent edge creation and deletion algorithms. I.e., we model processes where edge dynamics is defined as a function of time.

Our results provide a baseline for changes to be expected in dynamic networks. Also, they suggest that certain network properties have a strong, non-trivial dependence on the length of the sampling window.

**Keywords.** dynamic network analysis; cumulative network properties; time-dependent edge dynamics

## 1 Introduction

Social Network Analysis (SNA) is a field of research of growing importance. However, traditionally, it works with static networks (i.e., with systems where interaction topology is constant over time). This means, that the starting point of traditional SNA is either a snapshot of a network in a specific point of time, or some kind of aggregation (e.g., a collection of longitudinal samples of networks). While there are networks that usually do not change their topology over time (e.g., metabolic networks or blood vessels) and are fundamentally static, many of them change substantially over time, thus both of the above approaches may miss important facts and tendencies of these networks.

Recent papers seem to broaden the analysis of networks changing in time (i.e., dynamic networks) and studies in the subject getting more and more attention.

In this paper we extended our investigation with further elementary dynamic network models as an attempt to study the baseline properties of dynamic networks, with the intention to analyze the structural property changes induced by the effects of time-dependent edge dynamics. Our goal is to understand the inherent properties brought about by the changes to elementary idealized static networks through dynamic processes.

Focusing on edge dynamics, in earlier works [12], we defined three dynamic versions of Erdős-Rényi [4] models, one of which can also interpreted as a dynamic Watts-Strogatz [16] network, and a variety of dynamic Preferential Attachment [3] versions. In the present paper, we extend this line of research by introducing new elementary models that apply time-dependent edge dynamics.

Working with dynamic networks (either collecting longitudinal samples of networks or trying to model the evolution of networks in time), one realises that sampling always involves the act of aggregation (cumulating interactions that happened in a certain time window into one single static network instance). This problem is unavoidable, but poorly studied (See Figure 1.). Therefore, we pay special attention to the length of the cumulation time window and its effect to the aggregated network's properties like density, degree distribution, or clustering.

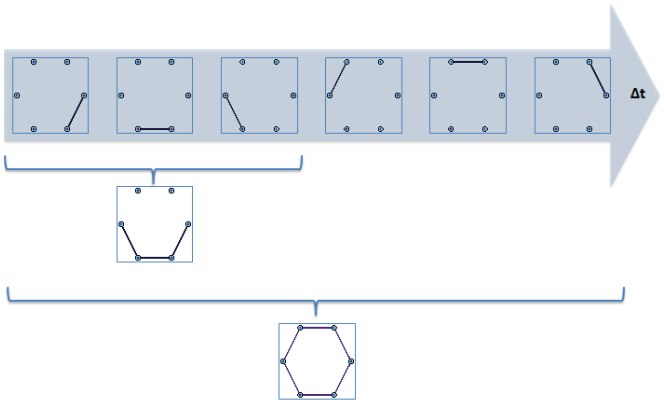



Figure 1: The importance of the sampling window's length

*Richard O. Legendi and Laszlo Gulyas are with AITIA International, Inc. E-mails: rlegendi@aitia.ai, lgulyas@aitia.ai

†Richard O. Legendi is with Eötvös Loránd University, Department of Programming Languages and Compilers, Faculty of Informatics, Hungary. E-mail: legendi@inf.elte.hu

‡Laszlo Gulyas is with Collegium Budapest, Hungary. E-mail: lgulyas@colbud.hu

Our work has a direct implication to the interpretation of SNA, because dissemination of information, innovation, or even epidemic processes may unfold inherently differently if we use different time windows of investigation.

The paper is structured as follows. The next section discusses basic considerations regarding dynamic networks and our research questions relating to them. Section 3 formalizes this by providing basic definitions and introduces the Elementary Dynamic Network models of our studies, and reports our results, discusses related work. The last section outlines directions for future work and concludes the paper.

# 2 Preliminaries

Dynamic processes in connection with social networks is a popular subject of investigation. This usually means two kind of studies. One class studies dynamics on networks, i.e., the cascading state changes of network nodes (percolation or contagion processes [14], information cascades [15], [1], etc.). The other class is concerned with the dynamics of the network: the structural, topological changes among the nodes and their links that emerge in time [10], [9], [2], [13]. This is the subject of our work in this paper. A third, smaller, class of works also exists that assume that these two processes are not independent: structural changes (i.e., new links) induce state changes and vice versa [6].

## 2.1 Taxonomy of Network Dynamics

Let us start with making a few elementary considerations. Given a network of $G(V, E)$, structural change means changes in the node set $V$ and/or changes in the links $E$. That is, the changes in the node set can be classified as:

A. the node set does not change,

B. a period when the node set remains constant,

C. the size of the node set increases, and

D. the size of the node set decreases

In real-life examples, the timescale on which new edges are added to the network is far shorter than how often new nodes are appended to the network (e.g., in case of social networks people tend to form and break up connections, which can be measured in hours or days, while it may take several years for new members to join or leave the network). For this reason, in our first part of our research we focus on constant node set as described by scenario A. We expect that the modification of the node set will not be a major factor determining the properties of social networks.

On the other hand, the number of links is a key property of networks, often expressed as the density (ratio of links per possible links), or as the average number of links per node. In this paper we will work with the former and classify link changing processes as those of

1. constant density,
2. increasing density, and
3. decreasing density.

Naturally, a dynamic process can show a mixture of the above basic scenarios. For example, for a period, it may display an activity that increases the number of nodes, while keeping density constant. Then it may continue by switching to a period when the nodeset remains constant but link density falls, and so forth. In the first part of our studies reported in this paper we will consider only *homogeneous* dynamics: working under the assumption that the different sections can be investigated in separation. As we will see later, certain processes induce dynamics where the measured property (i.e., density) is constant *on average*, while often diverging from the exact value dictated by this value. Therefore, we will allow for minor oscillations.

## 2.2 What to Study?

Many structural and statistical properties of networks are well known from traditional network analysis. In a dynamic setting, it is but natural to study their temporal evolution. Therefore, in this first analysis of elementary dynamic network models, we focus on the evolution of the number of components ($CN$), size of the largest component($S_{max}$), clustering ($C$), average degree ($D_{avg}$), degree distribution and most importantly, average path length ($l$) and density ($d$). The reason why $l$ and $d$ might be of special interest is the fact that there is a direct connection between them and the *average* of various centrality measures [7], [8].

- Average degree centrality is directly and trivially connected to density.
- The connection between the sum of betweenness centralities and the average path length: the average betweenness in a network will be about equal to the average path length in the network
- The connection between the sum of closeness centralities and the average path length

Betweenness centrality may also be relevant for several reasons, so we also included the average and maximum betweenness centrality to the list of evaluated metrics ($BC_{avg}$ and $BC_{max}$, respectively). The results then may be compared with the expectations based on the conclusions enumerated in the listing above (i.e., connection between the centralities and the average path length).

To compare the difference of degree distributions we used the Gini-coefficient [5], which shows the inequality of the distribution ranging from the value of 0 (*expressing total equality*) and the value of 1 (*representing maximal inequality*).

Another interesting aspect is to consider different aggregations of the individual network instances. Many properties of dynamic networks depend on the length of the time window within which network events (e.g., the appearance and disappearance of links) are aggregated. Therefore, we analyze the above measures both for the individual networks generated by the events and for the networks yielded by various time windows of cumulation. In fact, in many cases (when the observation time unit is fine enough) individual *snapshot* networks are not even interesting (i.e., they are very sparse, having almost no links, so it is not even connected, just a composed of a bunch of unconnected components). However, they become interesting when a number of them are *cumulated* together. Naturally, as the cumulation window length grows further, *cumulative networks* become less interesting as they will almost surely eventually approach the complete network (with all possible edges present). One of our main purpose with this research is to identify the time window ranges to use, where the resutlting network features meaningful properties.

The properties of the snapshot and cumulated networks are inherently different: the former is usually an unconnected graph with several components and isolated vertices; the latter is a growing graph that converges to the full graph as $t \to \infty$.

In the following section we describe the natural expectations one can think of about the models, and discuss the observed results.

### 2.3 Expectations

Based on the fact that the cumulative network accumulates all created edges, but never looses them, the following observations can be made:

1. density ($D$) grows with cumulation window length ($T$), and it has direct connection to other properties (e.g. centralities [7], [8])
2. Transitivity (measured by clustering, $C$) grows with density ($D$)
3. Average path length ($l$) drops with density ($D$)
4. Number of components ($CN$) drops with density ($D$)
5. Size of the largest component ($S_{max}$) grows with density ($D$)

The aforementioned simple expectations are obvious, however, the transition rate how the properties change is still a question that may show unexpected trends.

## 3 Formulation

Formally speaking, a dynamic network is a series of graphs. That is, $DN = G_t(V_t, E_t)$, where $E_t \subseteq V_t \times V_t$ ($\forall t \geq 0$). The initial network, $G_0$, is considered as a parameter of the process. In the current discussion, we assume that the node set is fixed. That is, $V_t = V_0$ for all $t \geq 0$. In addition, we assume that the evolution of the network (i.e., the changing composition of links) can be described as the result of the workings of two processes: i) an edge creation process and ii) an edge deletion process, that both identify a set of edges, $CE_t \subseteq (V_t \times V_t) \setminus E_t$ and $DE_t \subseteq E_t$, respectively. Given these, the edge dynamics of the network can be conceptualized as:

$$E_{t+1} = (E_t \setminus DE_t) \cup CE_t$$

Let us introduce the following notations. The number of nodes in the network will be denoted by $N = N_t$. The density of the network is the ratio of the number of existing edges over the maximum possible number of edges in the network of a given size. Assuming an undirected network where loops (self-links) and hyperlinks are not permitted, the density is given as

$$d_t = \frac{2\,|E_t|}{N(N-1)} \approx \frac{|E_t|}{N^2}$$

As we stated in section 2, we follow a *constant density assumption* during our current research, formally:

$$\exists \varepsilon > 0 : \forall t \in [0, T] : |d_t - d_{t-1}| \leq \varepsilon$$

### 3.1 Envelopes of Dynamic Networks

For any model of dynamic networks, there is a number of various networks of interest. The most immediate ones to consider are the momentary *snapshots* of the dynamic process, i.e., the network $G_t$ that exists at any given time $t$. (Conceptually, these may also be aggregations of individual network events, like changes in the relationship between a pair of nodes, but these events occur so close to each other in time that we cannot distinguish the time of their occurence.) In practice, however, we often deal with the aggregation of these networks. In the following, we study the properties of networks that are generated by the *cumulation* of a number of $G_t$'s. For reasons of simplicity, in the following we only deal with cumulative series starting at $t = 0$ (this can be done without the loss of generality). Thus, for the purposes of this paper, we define a *cumulative network* as

$$G_T = (\bigcup_{t=0}^{T} V_t, \bigcup_{t=0}^{T} E_t) \text{ for } T \geq 0.$$

### 3.2 Elementary Models of Dynamic Networks

In our previous studies [12] we created elementary models of dynamic networks based on the classic network models for static networks. The three classic models we had our eyes on were the Erdős-Rényi (ER) model of *random* density networks, the *small–world* network of Watts and Strogatz (WS) and the *preferential attachment* model of Albert and Barabási for *scale-free netwroks* (PA).

In our current study, we extended these models with two new ER models, where the presence of edges is directly connected to the simulation time. The following sections describe the evaluated models with the analysis and conjectures for the cumulative networks.

### 3.3 ER4

We start from an initial graph $G_0$ created by the static Erdős-Rényi model with density of $p_0$. Our previous approaches to edge addition and deletion processes was independent (e.g., random, preferentioal or based on a probability). In our current study, we extend we extend our investigations by time-based modification processes: we assing the *time of presence* of an edge by choosing a random number from the interval $[0, T]$ uniformly, where $T$ is the simulation time. The edge is present in the snapshot network only in the timestep of $t$ (cf. Lee et al. [11])). So basically it is a model where random links are created with a random entry time and lifetime.

The basic idea behind the model to get a general idea about what properties does the snapshot and cumulative network has when we know that we get an Erdős-Rényi model as an aggregated network.

One thing to note is that in our previous models, the density of the *snapshot* network was adjustable. For the ER4 model, the density means the density of the *prototype* network (which is equal to the cumulative network at the end of the simulation time, $T$). However, the expected density of the snapshot network can be derived by scaling it to the different $G_t$ instances: $d_{G_t} = d_0/T$.

#### 3.3.1 Experiments and Results

We performed the simulation runs with the parameter settings for 100 nodes, an initial density for the prototype network $d = 0.02$. We set the edge lifetime to $s = 5, 10, 15, 20, 25$, setting the final simulation time to $T = 100$. To validate the results, we performed 125 simulation steps in practice. We have two seed parameters for the random generators (initialization and addition seeds). We performed the whole simulation with 3 different values for each seed parameter (9 runs in total), the results here are the averaged values of the results we obtained.

**Density** Density shows monolithic increase with the same for the cumulative networks reaching very similar values in the end. That is no surprise, since the edges are distributed uniformly in the interval $[0, T]$. In the beginning, this trend is followed by the snapshot networks to the point when time reaches the lifetime of the edges (it is shown for the lifetime is 20 and 25 cases). The edge lifetime has a direct influence on the density level, as expected: as we increase the lifetime the density increases linearly for the snapshot networks. At the end all of the snapshot networks become empty networks. The average degree shows exactly the same trend (since its value can be derived from density directly).

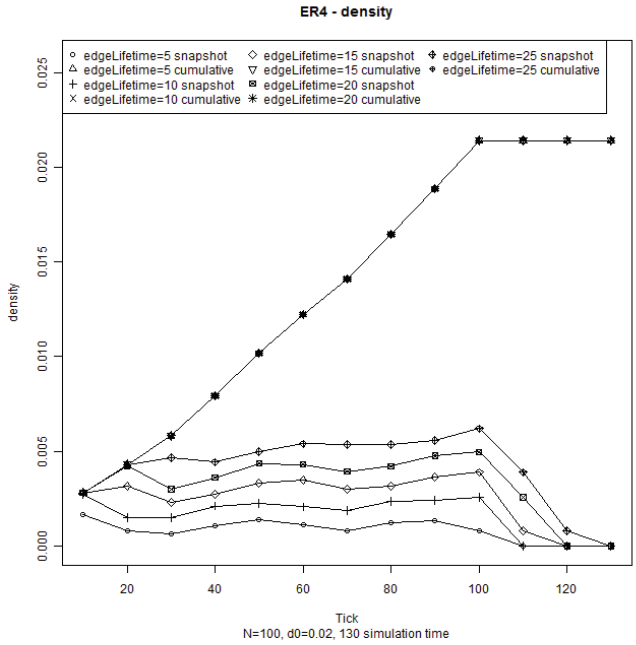



Figure 2: ER4 Density

**Betweenness** The betweenness values for the snapshot and cumulative network show significant differences. The average and maximum betwennes values show the same trends (but of course, on a different scale).

The betweenness value of the cumulative network shows some interesting phenomena to note. First, there is a very sharp transition between the 70th and 80th tick, the betweenness centrality grows by more than three times. Second, it is not a monotonic increasement, it continues increasing up to a point but then the average betweenness value falls back to lover rate (which then remains constant since there is no new edge added in the last 25 turns). That may happen in cases when new edges introduce several alternate paths between two components of the network.

The snapshot network shows only minimal differences, since the density of the network is very low as shown before. The maximum betweenness shows exactly the same trends as the average betweenness.

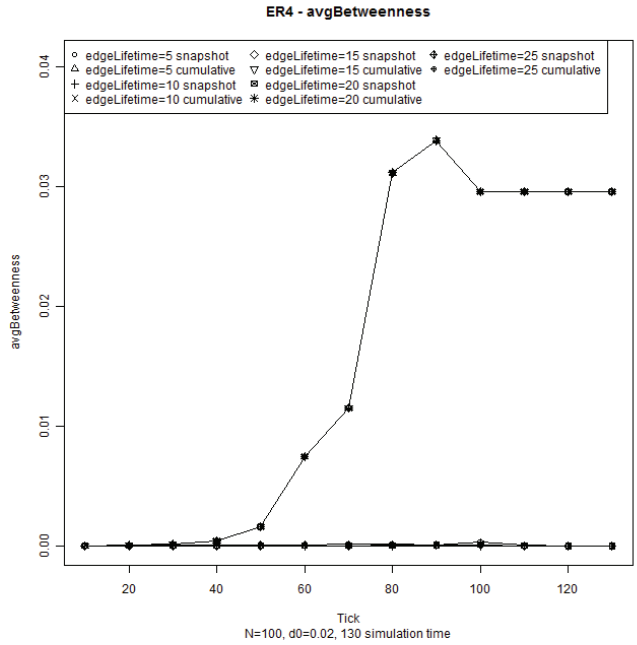



Figure 3: ER4 Average Betweenness Centrality

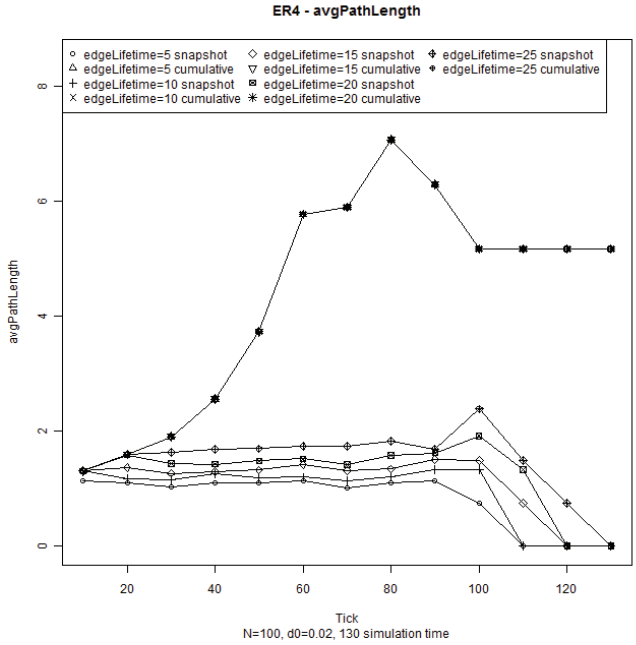



Figure 4: ER4 Average Path Length

**Average Path Length** The average path length reflects the same sort of change: the cumulative network has an outstanding value at the 80th tick, and starts decreasing.

**Clustering** The clustering shows intresting values as well. For the cumulative network, the values shows similiar trends to the average betweenness: a sharp transition may be observed a bit sooner in the clustering (between the 50th and 60th time step), and the fall back of the value may also be seen after 70th tick. Another interesting phenomenon here is that the clustering value of the snapshot network may exceed the clustering of the cumulative network by several times the in some of the cases. This situation have occured with even two of the models around the same time steps, these were the ones with the higher edge lifetime values (20 and 25).

**Components** The number of components show expected results. With the given number of nodes and initial density value, the fractional size of the maximum component of the cumulative network approaches a high value at the end of a simulation. The number of components has a linear connection to the edge lifetime, similarly to the density values.

As a general note, we can conclude that the results for the cumulative networks are almost identical.

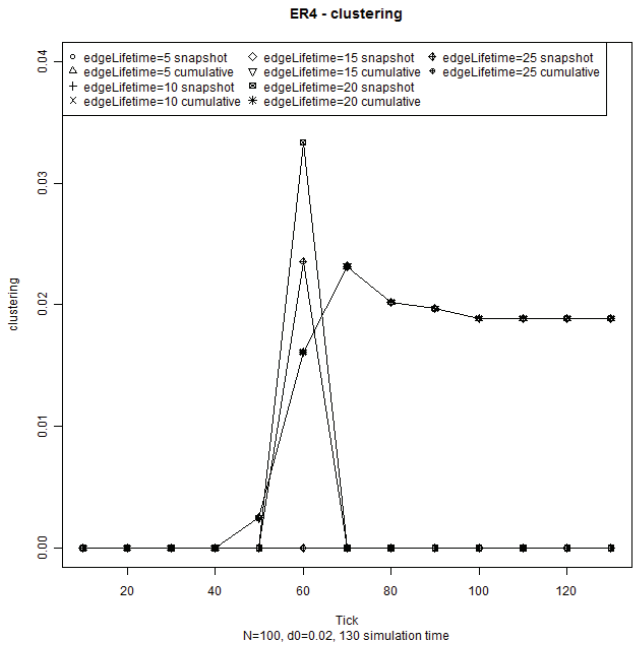



Figure 5: ER4 Clustering

## 3.4 ER5

This model is based on ER4, but with a minor modification. Edges in this model appear periodically in each

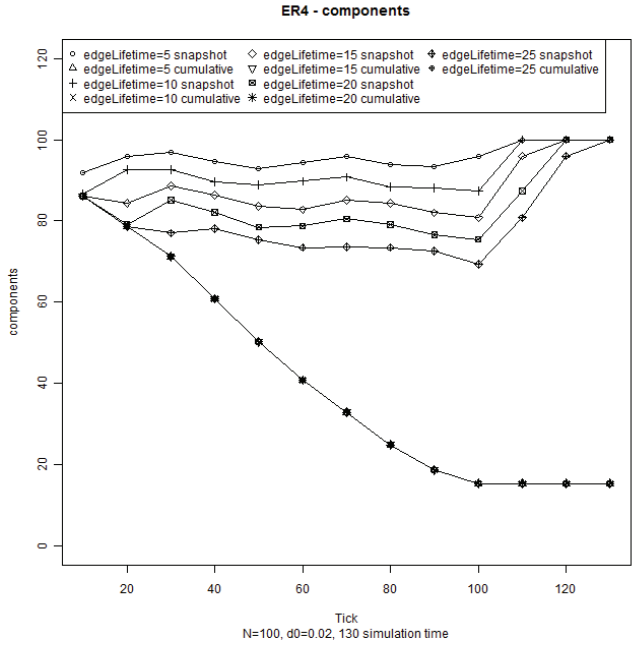



Figure 6: ER4 Components

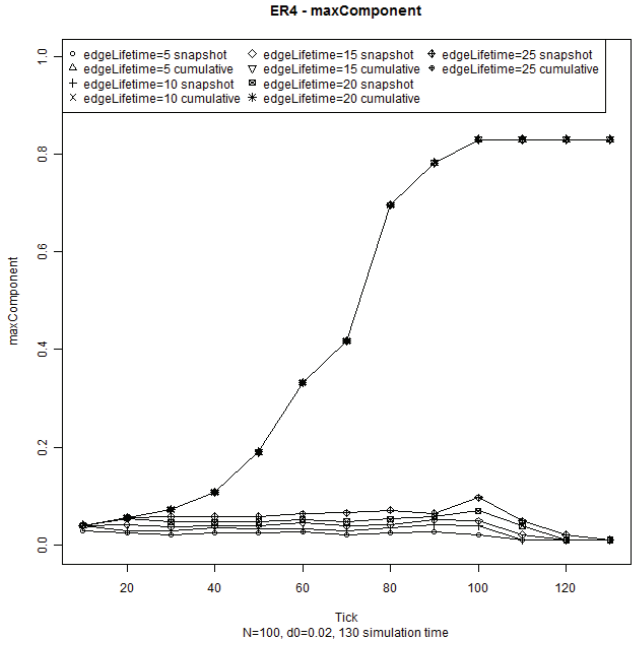



Figure 7: ER4 Fractional Size of the Maximal Component

$k * s$ timestep: random links are created with a random appearance frequency and lifetime (cf. Lee et al. [11]).

### 3.4.1 Experiments and Results

We performed the simulation runs with the parameter settings for 100 nodes, an initial density for the prototype network $d = 0.02$, setting the final simulation time to $T = 100$. We have two seed parameters for the random generators (for the initialization and addition). We performed the whole simulation with 3 different values for each seed parameter (9 runs in total), the summarized averaged values of the results we obtained may be found in Table 3.4.1.

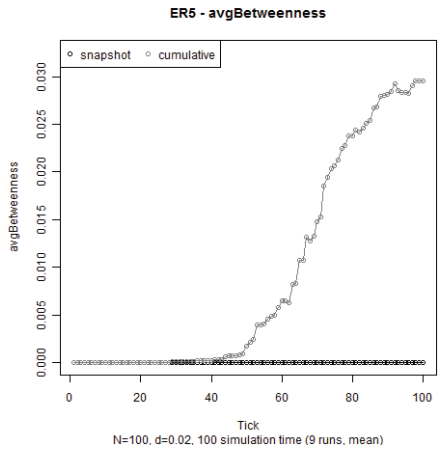
An interesting phenomena is that the cumulative network properties show smaller increasements up to the half of the simulation around the 50th tick, and shows significant increasements afterwards. The average and maximum betweenness centrality remains a very small value but starts growing in a very quick way. Density and the average degree grows in a linearly, it is because edges and their frequency is distributed uniformly between 0 and $T$. This is also the reason why the number of components decrease in a liear way as well. The average path length reflects a very sharp transition in the first few simulation steps, and starts growing to a level until the giant component appears in the network, and starts decreasing with new and new edges. The number of components and the fractional size of the maximal component show us the current size of the giant component.

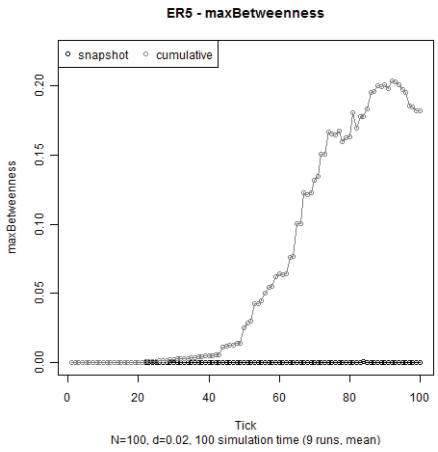
An extremely volatile and sensitive property for the cumulative network is the clustering which is also oscillating wildly. It has absolutely no meaning until the 50th simulation step, but increases within a few steps. Its value is not even stable, it may change within in every few steps.

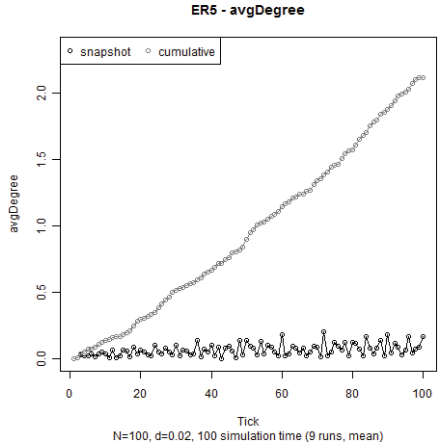
# 4 Conclusion and Future Work

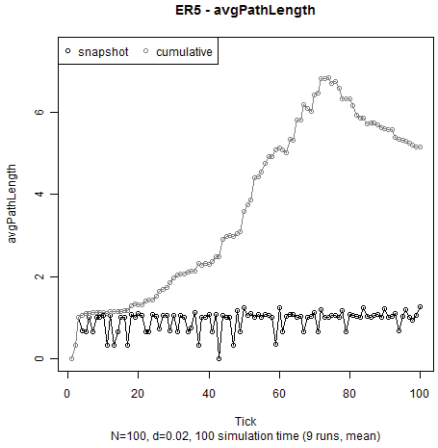
In this work we defined elementary dynamic network models as an attempt to study the baseline properties (number and size of components, clustering, degree distributions, average path length and density) of networks changing in time. Our goal was to understand the inherent properties brought about by the changes themselves to elementary idealized static networks through dynamic processes. Focusing on edge dynamics, we defined several versions Erdős-Rényi, Watts-Strogatz and a variety of dynamic Preferential Attachment versions.

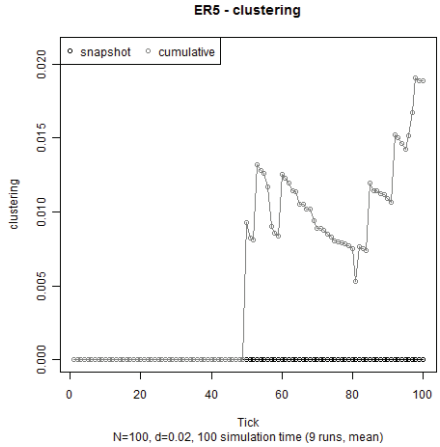
We found that, in case of these elementary models of dynamic networks, several network statistics are very sensitive to the length of the cumulation window. In particular, the density of the cumulative network changes superlinearly, especially in the relevant region of small to medium values. This is very important, since density has a great impact on the value of various other network measures.

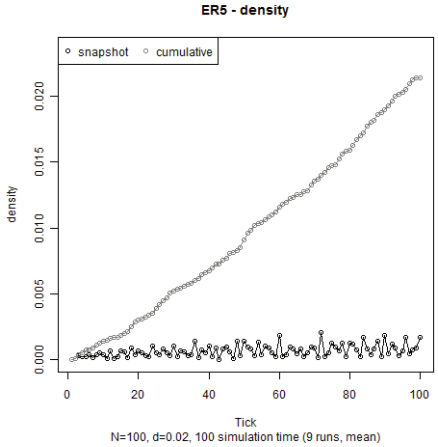

ER5 - avgBetweenness
snapshot
cumulative
avgBetweenness
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

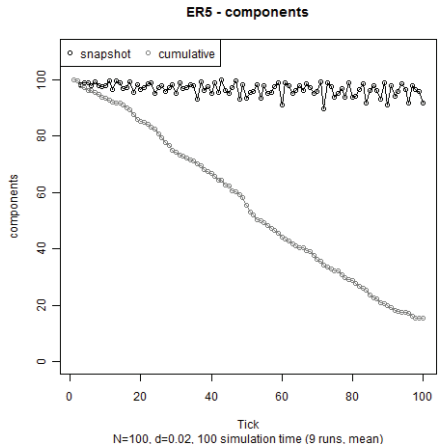

ER5 - maxBetweenness
snapshot
cumulative
maxBetweenness
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

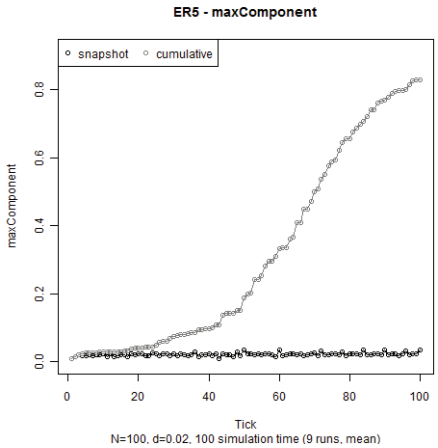

ER5 - avgDegree
snapshot
cumulative
avgDegree
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - avgPathLength
snapshot
cumulative
avgPathLength
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - clustering
snapshot
cumulative
clustering
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - density
snapshot
cumulative
density
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - components
snapshot
cumulative
components
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - maxComponent
snapshot
cumulative
maxComponent
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

The result that the length of the accumulation window makes a difference, seem to be trivial. On the other hand, this problem is not even mentioned in most of the works.

Our results also shed light on the behavior of the various models, with regard to how much history (i.e., initial network structure) matters. In some cases the network dynamics generated by these elementary models eradicate history instantly. In some other cases, history had an impact on both the snapshot and the cumulative network for an extended period of time.

### 4.1 Future Work

In future works we plan to extend our studies to include experiments with models that have richer and more realistic edge dynamics. These will include the following, still simple, models of elementary dynamic networks.

**Double Preferential** Choosing both ends of new links preferenitally.

**Assortative Mixing** Add links between a random node and another selected randomly with probability to the end node degree's similarity to that of the starting node.

**Lonely Node Preferentials** Only disconnected nodes create new links (approximation of newcomers in original PA).

To derive the formulas for the probability distributions of the metrics (or for the expected value of the metrics) along time would be another research direction, since this would allow for a precise comparison between the different models and networks, and for a precise comparison of the usage of different window lengths within the same dynamic model.

## Acknowledgements

This research was partially supported by the Hungarian Government (KMOP-1.1.2-08/1-2008-0002 ) and the European Union's Seventh Framework Programme: DynaNets, FET-Open project no. FET-233847 (`http://www.dynanets.org`). The supports are gratefully acknowledged.

## References

[1] Andrea Apolloni and Floriana Gargiulo. Diffusion processes through social groups' dynamics. *1103.0973*, March 2011.

[2] Andrea Apolloni, Achla Marathe, and Zhengzheng Pan. A longitudinal view of the relationship between social marginalization and obesity. In John Salerno, Shanchieh Jay Yang, Dana Nau, and Sun-Ki Chai, editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 6589, pages 61–68. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509 –512, October 1999.

[4] P. Erdős and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[5] C. W. Gini. Variability and mutability, contribution to the study of statistical distributions and relations. *Studi Economico-Giuridici della R. Universita de Cagliari*, 1912. Reviewed in: Light, R.J., Margolin, B.H.: An Analysis of Variance for Categorical Data - J. American Statistical Association, Vol. 66 pp. 534-544 (1971).

[6] Laszlo Gulyas and Elenna R Dugundji. Emergent opinion dynamics on endogenous networks. http://adsabs.harvard.edu/abs/2006physics..10125G, October 2006.

[7] Laszlo et al. Gulyas. Betweenness centrality as a function of network density and size. *(under publication)*.

[8] László Gulyás, Gábor Horváth, Tamás Cséri, Zalán Szakolczi, and George Kampis. Betweenness centrality dynamics in networks of changing density. *Presented at the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010)*, 2010.

[9] L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Rava , C. Rizzo, and A. E Tozzi. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *1104.2515*, April 2011. PLoS ONE 6(2): e17144 (2011).

[10] Lorenzo Isella, Juliette Stehle, Alain Barrat, Ciro Cattuto, Jean-Francois Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *1006.1260*, June 2010. J. Theor. Biol. 271 (2011) 166-180.

[11] Sungmin Lee, Luis E. C Rocha, Fredrik Liljeros, and Petter Holme. Exploiting temporal network structures of human interaction to effectively immunize populations. *1011.3928*, November 2010.

[12] Richard O. Legendi and Laszlo Gulyas. Cumulative properties of dynamic social networks. 2011. Under publication.

[13] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1, March 2007. ACM ID: 1217301.

[14] Juliette Stehle, Alain Barrat, and Ginestra Bianconi. Dynamical and bursty interactions in social networks. *1002.4109*, February 2010. Phys. Rev. E 81, 035101(R) (2010).

[15] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766 –5771, April 2002.

[16] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, June 1998.