Chapter 1

Driving signals

The complex networks grow by adding new nodes, and growing network models consider growth constant over time. This approximation is sufficient for explaining how properties of complex networks can emerge; for example, we find scaling of degree distribution in the Barabasi-Albert model, such as in real systems. Models mainly focus on linking rules and their influence on the topology of complex networks.

Still, the growth of real systems changes over time. In online social networks, new users join daily, and the users' activity might have bursty nature. We can consider a co-authorship network, where links are created between scientists when they publish a paper. The dynamic of real networks can be complex and highly influenced by non-linear signals. The growth signal, the number of new nodes in each time step, has cycles and trends. Circadian cycles are directly reflected in growth signals, and we also find long-range correlations and multifractal properties.

In this chapter, we explain the properties of growth signals, both real and computer-generated. We analyze networks created with a growing network model where the interplay between ageing and preferential attachment shapes their structure. We are interested in incorporating non-constant growth signals into the model and measuring their impact on the complex networks. Differences between networks with the same number of nodes and links can be observed by analyzing connectivity patterns. Figure 1.1 describes our goals.

1.1 Growing network model with growing signal

To enable nonlinear network growth in the number of nodes, we need to adapt the existing models such that at each time step, we can add $M \geq 1$ new nodes that make $L \geq 1$ links with existing nodes in the network. The master equation N_k, k degree nodes can be written as:

$$\partial_t N_k = \sum_{j=1}^{M(t)} r_{k-j \longrightarrow k} N_{k-j} - \sum_{j=1}^{M(t)} r_{k \longrightarrow k+j} N_k + M(t)\delta_{k,L}. \tag{1.1}$$

At each time step we add $M(t)$ nodes with L links. As multiply links between two nodes are not allowed, we'll get $M(t)$ new nodes with degree L, which describes the third term in the equation. Old nodes can increase their degree from 1 to $M(t)$, as different new nodes can choose the same node.

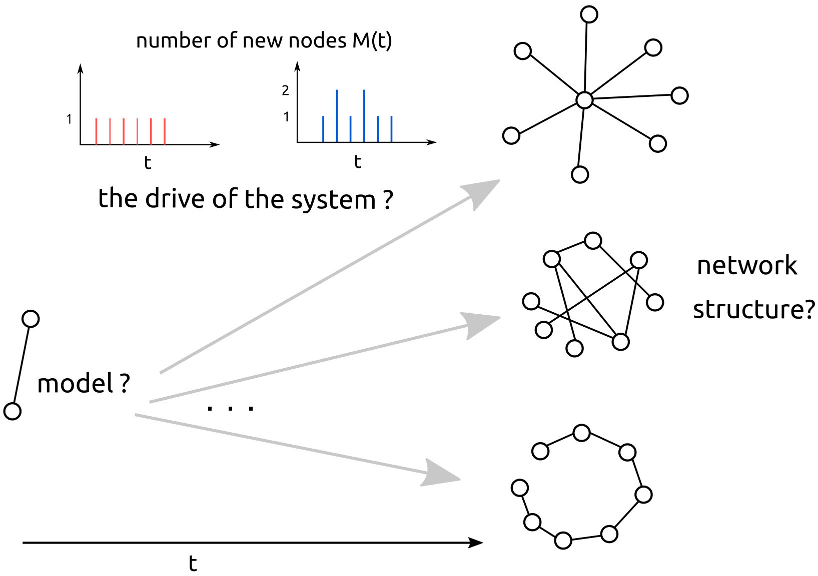

Figure 1.1: The open question is how nonlinear signals in combination with network model influence the structure of the network. Under what circumstances networks have the scale-free, hub-spoke or chain structure.

The first term in the equation describes nodes with degree $k \in \{k - M(t), \dots, k - 1\}$ that getting degree k, while in second term nodes with degree k entering degree $k \in \{k + 1, \dots, k + M(t)\}$. The quantities $r_{k-j \longrightarrow k}$ and $r_{k \longrightarrow k+j}$ are the rates that express the transitions of a node from class with degree $k - j$ to one with degree k and from class with degree k to class with degree $k + j$ respectively.

For model we choose, aging model where linking probability depends on network degree k and its age τ, $\Pi_i(t) \sim k_i(t)^\beta \tau_i^\alpha$. With this linking probability the master equation was solved for $M(t) = const. = 1$, using approach [1]. When $M(t)$ is the correlated function, the equation is not solvable analytically. Instead, we use numerical simulations to study the influence of the signal $M(t)$ on the network structure. When we add only one link per node $L = 1$, networks are uncorrelated trees. To obtain the clustered structures, we need to use $L > 1$; so each new node can create more than one link. Finally we focus on the aging model parameters $-\infty < \alpha \leq -1$ and $\beta \geq 1$. We expect critical line $\beta(\alpha^*)$ where scale-free networks can be found. Under critical line, networks have stretched exponential degree distribution, and for large β small-world networks are present.

Finally, we need to define what the time series of new nodes are. We focus on the growth of two real systems, the **TECH** [2] community in the Meetup website and on two months of **MySpace** [3] social network.

1.1.1 Time-series from real systems

MySpace signal is the number of new members who appear for the first time in the data. The dynamics of the Meetup website happen on different scales than on Meetup. Here, the time step is one minute. The MySpace signal has $T = 3162$ steps, with $N = 10000$ members. To describe the properties of the signal, we use Multifractal detrended analysis and calculate the Hurst exponent on different scales, showing the right pane of the figure, 1.2. It is multifractal $q < 0$, and becomes constant for $q > 0$, it has long-range correlations as $H(q = 2) = 0.6$. My Space signal has cycles characteristic of the human circadian rhythm, figure 1.2. If we randomize the MySpace signal, we find that we can easily destroy trends and cycles. The randomization is done with the reshuffling procedure, where we keep number where we keep the number of nodes, length and the mean value of the signal. The inset of the original and randomized signals show the time series' global profile; we find that trends are destroyed. Also, randomized MySpace signal does not have long-range correlations anymore, Hurst exponent indicates short-range correlations $H = 0.5$, and the signal becomes monofractal.

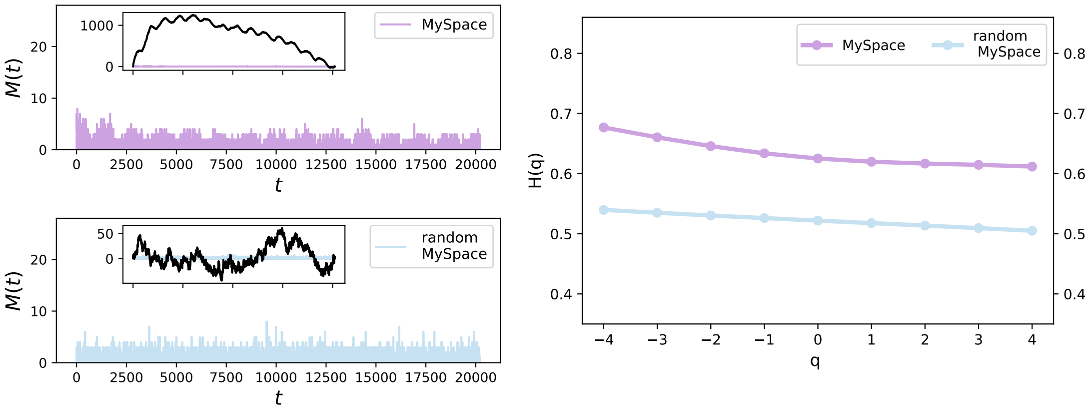

Figure 1.2: MySpace signal, the random MySpace signal (left pane) and the dependence of multi fractal Hurst exponent $H(q)$ of the scale q. (right pane)

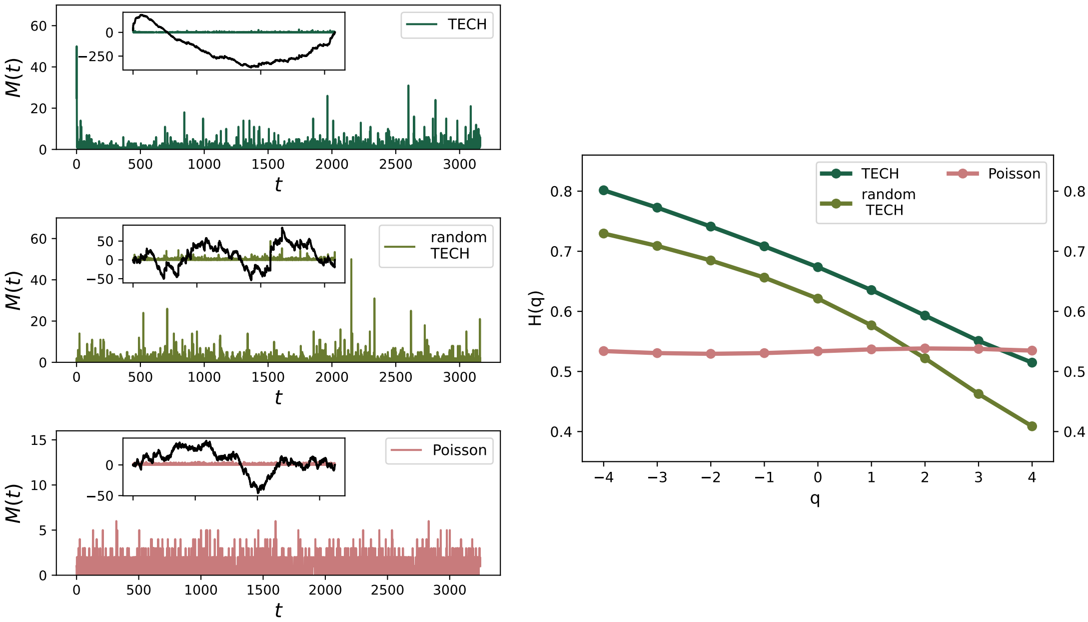

Figure 1.3: TECH signal, the random TECH signal (left pane) and the dependence of multi fractal Hurst exponent $H(q)$ of the scale q. (right pane)

The TECH is a group from the Meetup website that gathers users interested in technology. Using the Meetup website, they organize offline events. The time unit in this time series is an event since then are created links between event. The TECH time series M(t) represents the number of users who joined the TECH community and visited the event for the first time. The time series length is $T = 3162$ steps, and we count $N = 3217$ members in the TECH community for a given period, 1.3. TECH signal has long-range correlations with Hurst exponent $H(q = 2) = 0.6$. Also, we find that TECH is multifractal, as the Hurst exponent is not constant across the scales. The multifractality originates not only from signal trends but also from the broad probability distribution of time series. If we randomize the TECH signal, we find that we can easily destroy trends and cycles, but signal keeps multifractal properties, meaning that broad probability distribution can not be eliminated. For that reason, we generate the uncorrelated signal from Poissonian probability distribution. The length of this signal is $T = 3246$, while we keep the number of nodes N same as in TECH signal.

1.1.2 D-measure

We can compare the networks with the same number of nodes and links generated with growth signals with different properties. We use growing network model where we vary parameters $-3 < \alpha \leq -1$ and $-3 \leq \beta \leq 1$. We also vary the network density, $L \in \{1, 2, 3\}$. For each set of model parameters α, β, L and each signal $M(t)$, we create the sample of 100 networks. Besides this, for the same set of parameters, we generate the sample of networks with $N = 10000$ and $N = 3217$ nodes grown with constant signal $M(t) = 1$; one node is added to the network at each time step. To examine how different growing signals influence the structure of networks, we use D-measure [4], defined methodology chapter. We equally consider the global and local properties, setting parameter $w = 0.5$. We compare the networks grown with the constant and fluctuating signal with D-measure, for all network pairs between two samples, and averaging the result. The advantage this measure has is that it can measure the distance between two network structures, even if they are generated with the same model; that was not the case with Hamming distance or graph editing distance [4].

Obtained results for D-measure are presented on the figure 1.4. We note that the largest distance between networks is along critical line $\beta(\alpha^*)$ of the aging model. The fluctuations present in the signal mostly influence, the scale-free networks. For networks, away from this line, structural differences exist, but they are much smaller. For gel small world networks, $\beta > \beta^*$, the D-measure is close to zero. Under critical line, $\beta < \beta^*$, D-measure depends on the properties of the signal. If we fix network density L, position of critical line is independent of the properties of the signal. Still, with higher link density, critical line slightly move toward larger β, see figure 1.4.

In the region around the critical line, we find that D-measure depends on the properties of the signal. Multifractal signals TECH has the most considerable impact on network structure; the maximum value of the D-measure is $D_{max} = 0.552$. Similar behaviour is discovered for other multifractal signals, random TECH and MySpace. For networks generated with uncorrelated signals: random MySpace and Poisson, difference exists but it is much smaller.

D-measure rises for lower α. In the case of a constant signal, the number of nodes added to the network is equal for each time step, so at time interval T, the network has MT nodes. In fluctuating signal, the number of nodes added during time interval T vary. In signals, such as TECH, where there are peaks in the number of new users, hubs emerge faster. As we decrease the parameter α, fluctuations in the signal become more critical, and the hubs emerge even for uncorrelated signals. The trends in the real signals further promote the emergence of hubs in the network.

1.1.3 The structure of networks

We examine degree distribution, degree correlations and clustering coefficient of networks generated by real signals. It has been shown that these measures provide a sufficient set for describing the structure of complex networks. Results showed that multifractals influence networks more than monofractals; it is most prominent in scale-free networks.

Figure 1.5 shows properties of networks generated with model parameters $L = 2$, $\alpha = -1.0$, $\beta = 1.5$, that lies on critical line. The degree distributions $P(k)$ of networks generated with real signals TECH and MySpace have super-hubs emerged. Degree distributions generated with randomized and white noise signals do not differ from the degree distribution of networks generated with the constant

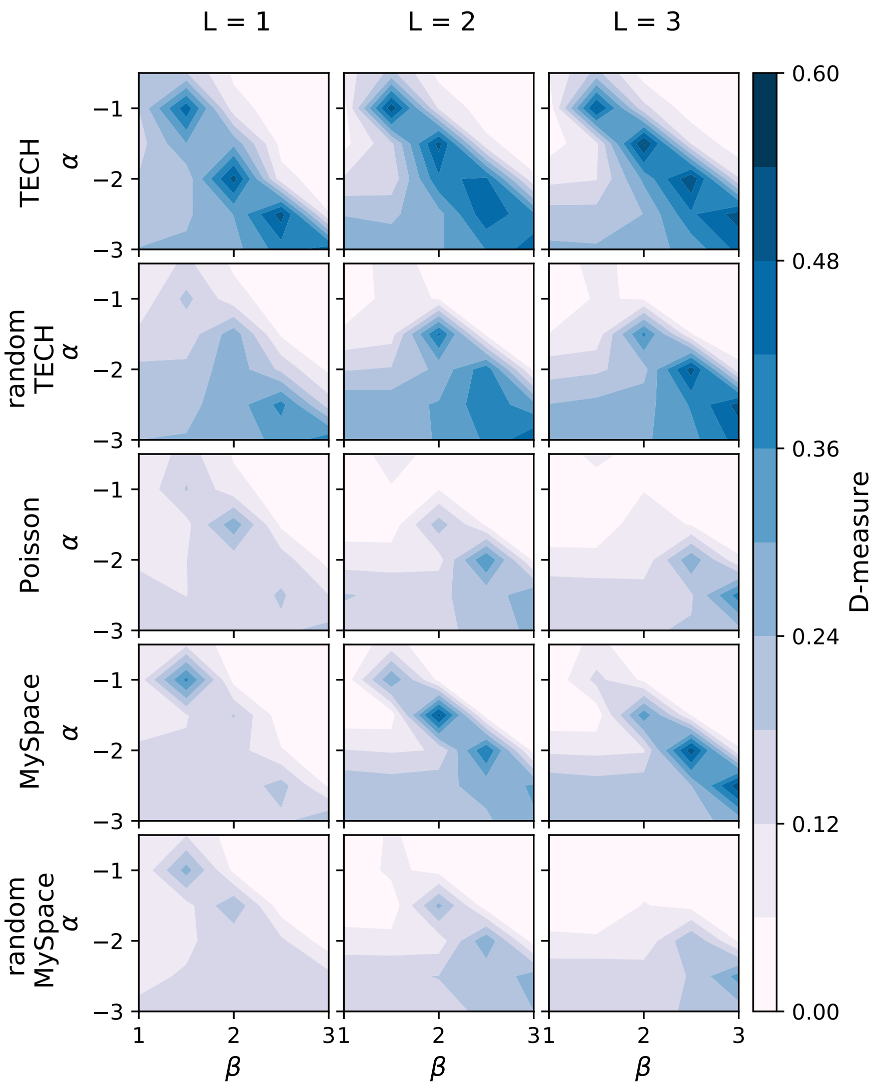

Figure 1.4: The comparison of networks grown with growth signals shown in figures 1.3 and 1.2 versus ones grown with constant signal $M = 1$, for value of parameter $\alpha \in [-3, -1]$ and $\beta \in [1, 3]$. $M(t)$ is the number of new nodes, and L is the number of links added to the network in each time step. The compared networks are of the same size.

signal. Networks generated with real signals average neighbouring degree $\langle k \rangle_{nn}(k)$ and clustering coefficient $c(k)$ depend on node degree, while in networks generated with constant and randomized signals, they weakly depend on the degree k.

We also find structural differences between networks, obtained with model parameters under the critical line $\alpha < \alpha^*$, see Figure 1.5. The difference is mostly found for TECH signal. Degree distribution $P(k)$ shows the emergence of hubs in networks grown with TECH signal, while the randomized and Poisson signals are more similar to networks grown with the constant signal. MySpace signal, whose generalized Hurst exponent $H(q)$ weakly depends on scale parameter q and whose long-range correlations and trends are easily destroyed, do not influence the structure of networks more than constant or randomized signal.

The properties of the time-varying signal do not influence the topological properties of small-world gel networks, Figure 1.5. Here model promotes the existence of hubs. As this is the mechanism through which the fluctuations alter the structure of evolving networks, the properties of the signal are not relevant.

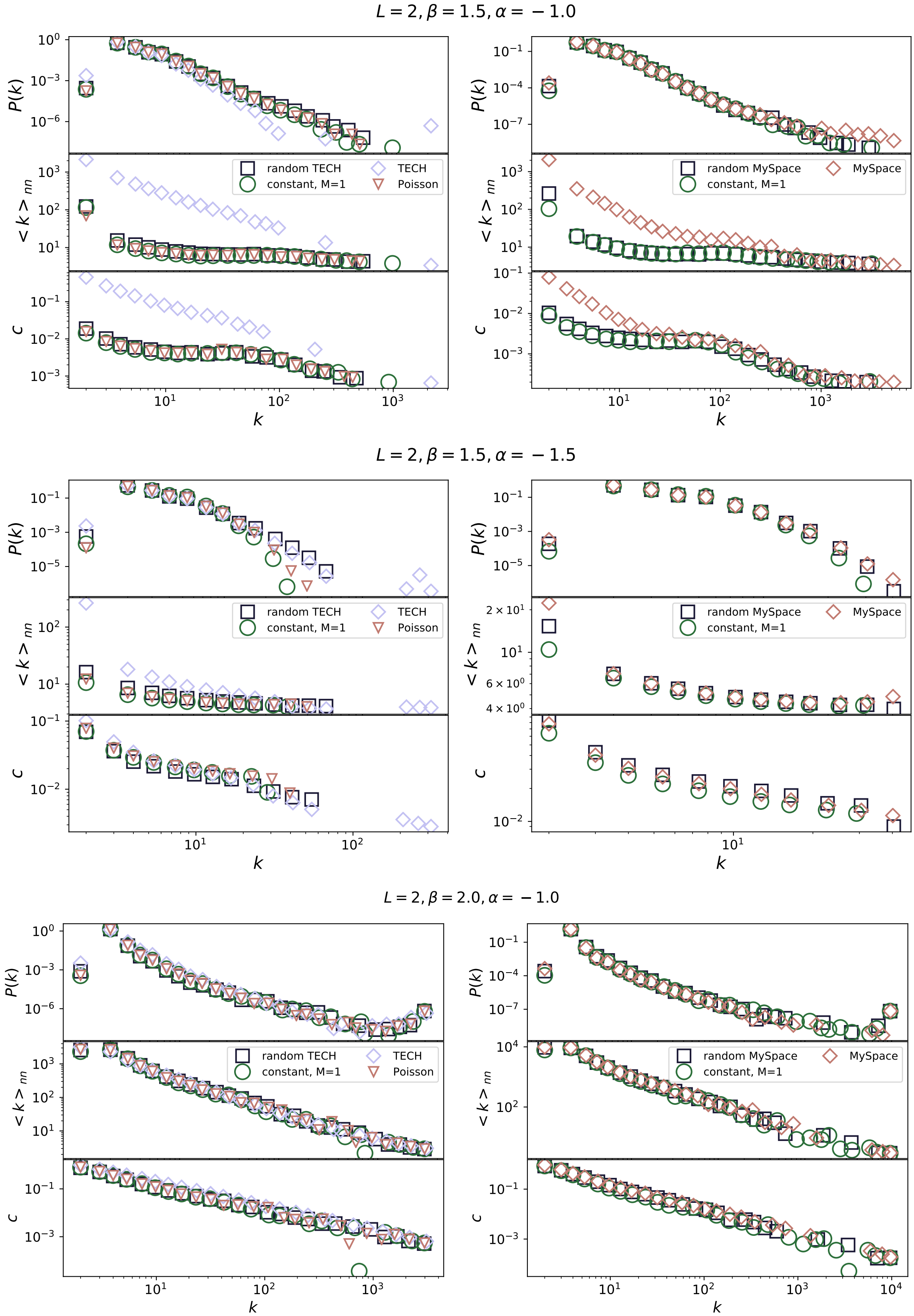

Figure 1.5: Degree distribution, the dependence of average first neighbor degree on node degree, dependence of node clustering on node degree for networks grown with different time-varying and constant signals. Model parameters have value $\alpha = -1.0$, $\beta = 1.5$ and $L = 2$ for all networks. The networks are from scale-free class. Model parameters have value $L = 2, \alpha = -1.5$, $\beta = 1.5$. The networks have stretched exponential degree distribution. Model parameters have value $L = 2, \alpha = -1.0$, $\beta = 2.0$. Generated networks have small-world properties.

1.2 Long range correlated signals

The previous section showed that the growth signal of real systems has complex dynamics. Besides long-range correlations, we also find multifractal properties, and it is hard to isolate individual effects and analyse their influence separately from each other. When this is the case, synthetic signals with specific characteristics help verify our findings in real systems. The long-range correlated properties can be included into time series using Fourier filtering transform method [5].

The data have power-law correlations $C(s) =< x_i x_{i+s} >= s^{-\gamma}$ characterized with coefficient γ. Hurst exponent depends on γ as $H = 1 - \frac{\gamma}{2}$. The Fourier transform, gives us the power spectrum of the time series $S(f)$, that is function of the frequency f. For the long-range correlated data it depends on coefficient $\beta = 1 - \gamma$ and has form:

$$S(f) \sim f^{-\beta} \tag{1.2}$$

We can generate the data using Fourier filtering with $\beta = 2H - 1$, as following:

- first generate one-dimensional sequence of uncorrelated random numbers u_i from Gaussian distribution with $\sigma = 1$
- calculate the Fourier transform of the generated sequence, u_q, the spectrum is flat as data correspond to white-noise
- then filter the power spectrum with $f^{-\beta/2}$, so the function will follow power spectrum expected for data with long-range correlations.
- calculate the inverse Fourier transform x_i. It transforms data to the time domain where signal has desired long range correlations

The Fourier filtering method gives the Gaussian distributed data, so data are without broad distributions, nonlinear or multifractal properties. Using this method we generated the signals for differet values of the Hurst exponent, see figure 1.6. The obtained signals are round to integers and mean values of signals are close to 4.

As before, we focus on the region of model phase diagram with negative α and positive β as there is found the transition line from stretched-exponential across scale-free to the small world- gel networks. We take range of parameters $-3 \leq \alpha \leq -0.5$ and $1 \leq \beta \leq 3$ with steps 0.5 and we also vary the the number of links each new node can create $L \in 1, 2, 3$. For each combination of (α, β, L) we generate the sample of 100 networks, and compare the structure of network grown with fluctuating signals with different Hurst exponent $H \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ and constant signal $M = 4$. The results represented by D-measure, shown in the figure 1.7 are obtained averaging the D-measure between all possible pairs of generated networks.

The higher values of D-measure are found in the region of critical line $\beta(\alpha^*)$. The most considerable influence is on networks with scale-free distribution. Comparing D-distance in only one point of phase diagram, for example $L = 1, \alpha = -2.5, \beta = 2.5$, we find correlations in the signal (Hurst exponent is larger), make bigger impact on the network structure. D-measure between networks grown with signal with Hurst exponent $H = 1.0$ and constant signal is $D(H = 1.0, M = 4) = 0.405$, while between networks grown with signal with $H = 0.8$ and constant signal is $D(H = 0.8, M = 4) = 0.316$. For $\alpha > \alpha^*$ networks have similar structural properties and D-measure is close to 0. In the region of networks with stretched exponential degree distribution $\alpha < \alpha^*$ differences are small.

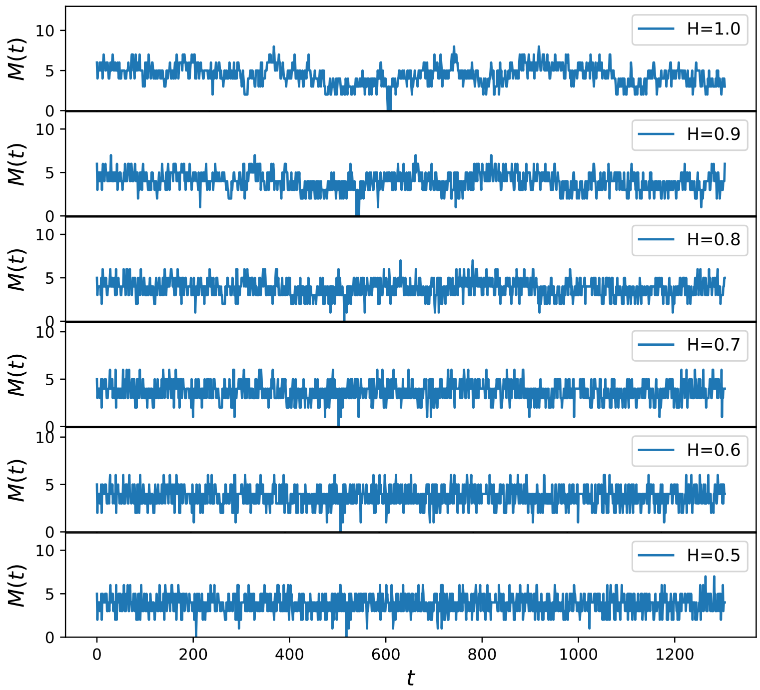

Figure 1.6: Monofractal signals

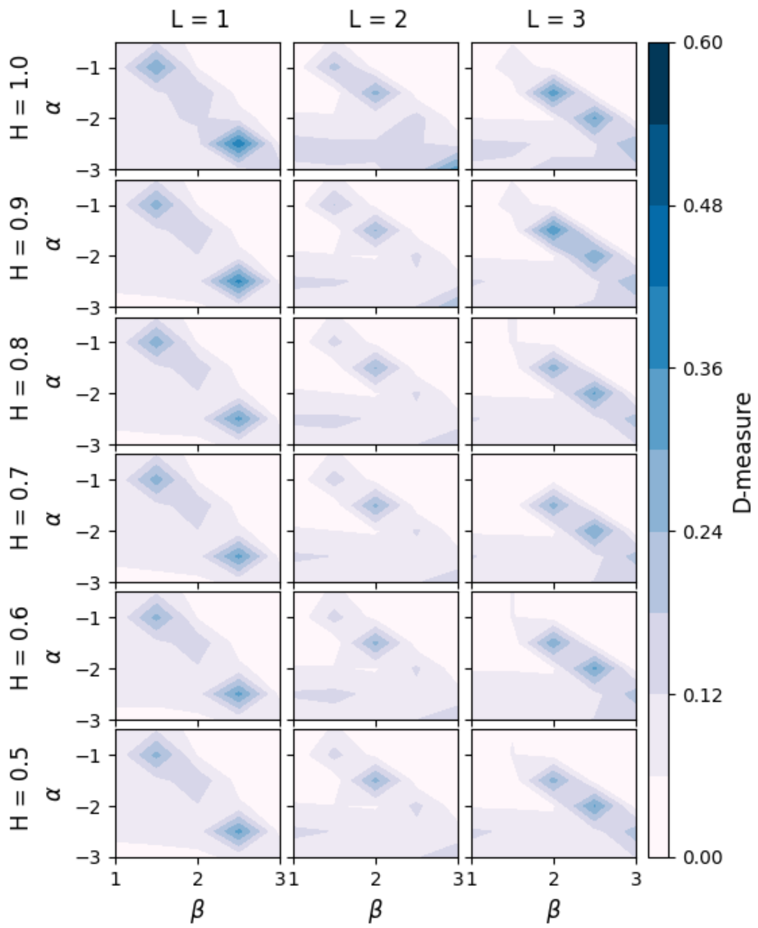

Figure 1.7: D-distance between networks generated with different long-range correlated signals with fixed value of Hurst exponent and networks generated with constant signal M=4.

We further explore the assortativity index and clustering coefficient of generated networks. We show results for several ageing model parameters to show the difference between networks this model can produce, 1.8. All networks are disassortative, with a negative degree-degree correlation index. For the values of parameters below critical line, $\alpha = -2.5, \beta = 1.5$ r does not depend on the Hurst exponent. Above the critical line are small-world networks. They are disassortative. The minimum value of the assortativity index is $r = -1$, for $L = 1$, indicating the presence of hubs connecting many nodes. The assortativity index slightly grows with link density.

In the region of critical parameters, the assortativity index depends on the value of the Hurst exponent. The larger influence on the assortativity index have correlated signals, with Hurst exponent $H > 0.8$, so networks become more disassortative, see line for parameters $L = 1, \alpha = -2.5, \beta = 2.5$ in Figure 1.8. The long-range correlations have a stronger effect on the evolution of networks with lower density.

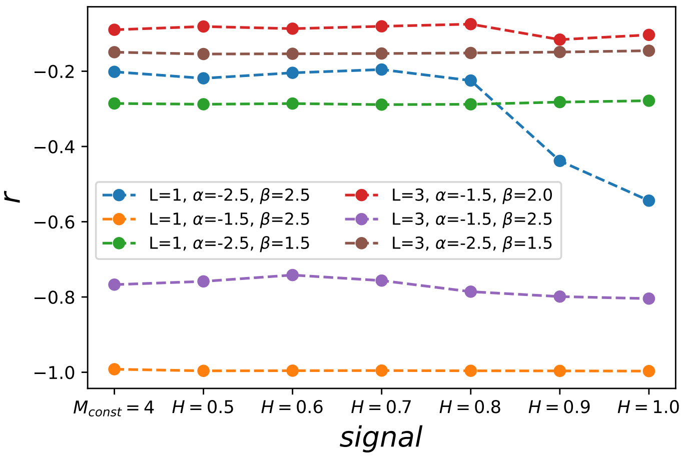

Figure 1.8: Mean assortativity index for networks generated with different with different model parameters α, β, L and different long-range correlated signals with Hurst exponent H.

Figure 1.8 shows the mean clustering coefficient. For $L = 1$, networks are uncorrelated trees, with clustering coefficient 0. For network density $L > 1$, nodes are organized into clusters. Under the critical line, for parameter $L = 3, \alpha = -2.5, \beta = 1.5$, clustering coefficient is constant and low. Similar values are obtained for clustering coefficient for critical parameters $L = 3, \alpha = -1.5, \beta = 2.0$, but for Hurst exponent $H > 0.8$ clustering coefficient increase. Small world networks, $L = 3, \alpha = -1.5, \beta = 2.5$ are clustered, the value of $< c >$ is high. The value of clustering for networks created with the constant signal is 0.8. Networks grown with white noise signal and signal with H=0.6 have higher clustering values, while networks grown with signals with a Hurst exponent larger than 0.6 have the same clustering value, below 0.8.

1.3 Conclusions

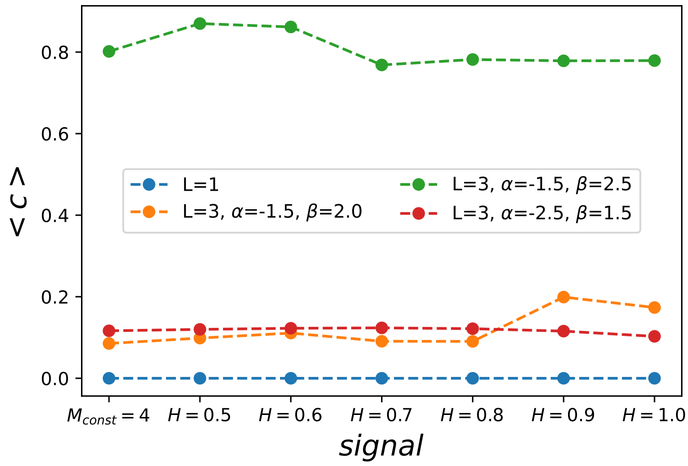

Figure 1.9: Clustering coefficient

Bibliography

[1] Sergey N Dorogovtsev and José FF Mendes. Scaling properties of scale-free evolving networks: Continuous approach. *Physical Review E*, 63(5):056125, 2001.

[2] Jelena Smiljanić and Marija Mitrović Dankulov. Associative nature of event participation dynamics: A network theory approach. *PloS one*, 12(2):e0171565, 2017.

[3] Milovan Suvakov, Marija Mitrovic, Vladimir Gligorijevic, and Bosiljka Tadic. How the online social networks are used: dialogues-based structure of myspace. *Journal of The Royal Society Interface*, 10(79):20120819, 2013.

[4] Tiago A Schieber, Laura Carpi, Albert Díaz-Guilera, Panos M Pardalos, Cristina Masoller, and Martín G Ravetti. Quantification of network structural dissimilarities. *Nature communications*, 8(1):1–10, 2017.

[5] Hernán A Makse, Shlomo Havlin, Moshe Schwartz, and H Eugene Stanley. Method for generating long-range correlations for large systems. *Physical Review E*, 53(5):5445, 1996.

[6] Marija Mitrović and Bosiljka Tadić. Emergence and structure of cybercommunities. In *Springer Optimization and Its Applications*, volume 57, pages 209–227. Springer International Publishing, 2012.

[7] Marija Mitrović Dankulov, Roderick Melnik, and Bosiljka Tadić. The dynamics of meaningful social interactions and the emergence of collective knowledge. *Scientific reports*, 5(1):1–10, 2015.

[8] Kamalika Basu Hajra and Parongama Sen. Phase transitions in an aging network. *Physical Review E*, 70(5):056103, 2004.

[9] Petter Holme and Jari äki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.