Examining Higher Order Transformations for Scale-free Small World Graphs

Chris Biemann
University of Leipzig, Germany
Email: biem@informatik.uni-leipzig.de

Uwe Quasthoff
University of Leipzig, Germany
Email: quasthoff@informatik.uni-leipzig.de

***Abstract*— The degree distribution of scale-free Small World networks follows a power law. For random graph generators, its exponent is constrained by the construction mechanism, whereas in real-world data, different slopes can be observed. However, the degree distribution alone does not reveal much of the local structure of these graphs. Therefore, we propose a graph transformation we call "higher order" transformation, which encodes the number of common neighbours two vertices share in its edge weights. Studying the degree distribution of second- and third order graphs and comparing it to natural language co-occurrence data, we find that the higher order transformation reveals differences that cannot be detected by only looking at traditional measures on the original graph.**

I. INTRODUCTION

Recent years showed an increased interest in random graph generators that yield graphs with the scale-free Small World property. The Small World (SW) property as defined in [6] is given by a higher average clustering coefficient and a similar average path length than in classical random graphs [4]. A graph is called scale-free when its degree distribution follows a power-law, i.e. there is no characteristic scale on the degree of vertices.

There is the following duality between SW graphs and vertex sequences: Using the trivial fact that every path in a graph corresponds to a vertex sequence, special vertex sequences are used to construct random SW graphs. These are defined by preference, a memory of the recently visited vertices and some random elements. These three features are essential ingredients of natural language, and SW graphs have been found in language data [5]. The aim of the paper is to introduce a higher order transformation for a comparison of these graphs.

A. Random Graph Models

Several models were proposed that randomly generate undirected scale-free Small World graphs by iteratively adding new vertices. In the BA-model [1], a graph is constructed by preferential attachment: a new vertex connects to existing vertices with a probability according to their degree. A modification that targets modelling natural language co-occurrences is the DM-model [2]: the new vertex is connected to the preferential existing vertex, but additionally edges in the set of existing vertices are introduced with a probability according to the product of their degrees.

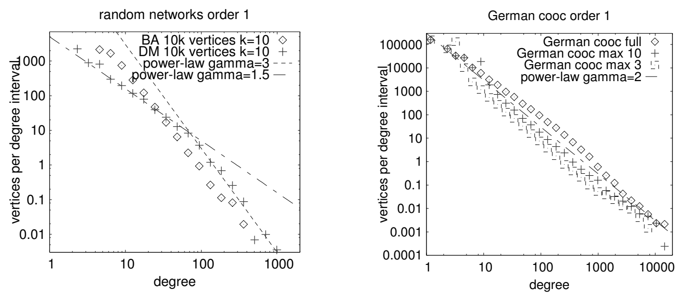

Fig. 1. Degree distributions (left) of two random graph models. (right) of significant sentence-based word co-occurrences of 1 million German sentences obtained from `http://corpora.informatik.uni-leipzig.de/download.html`. Data points were obtained using logarithmic binning. As other experiments showed, the characteristics are similar for all natural languages

In this work, we are merely interested in the degree distribution $D(k)$ of such random graphs. Figure 1 shows the degree distribution of graphs generated with the aforementioned models with $n = 10,000$ vertices and an average degree $k = 10$. $D_{BA}(k)$ is approximated by a power-law $pl(x) \sim x^{-\gamma}$ with $\gamma = 3$, $D_{DM}(k)$ follows two power-law regimes with $\gamma_1 = 1.5$ and $\gamma_2 = 3$.

B. Sentence Paths and Word Co-occurrences

Assume we are given a large but finite set of vertices V. A set of paths $P = \{P_1, ..., P_n\}$ (i.e. each P_i is a vertex sequence) defines a graph $G = G(V, P)$ having just the edges given in the paths, with graph characteristics depending on P. For the following, we assume a maximum path length l, which corresponds to the sentence length when representing words as vertices and paths as sentences in a text corpus. This procedure can be generalised by considering arbitrary graphs instead of paths. In the following, we will be interested in P consisting of relatively small connected subgraphs of small diameter and l or less vertices.

Considering weighted edges, we count how many elements of P contain the respective edge. Edge weights are relevant for the following *co-occurrence* graph. For this, we replace P_ks by P'_ks, where P'_k is the fully connected graph of vertices in P_k. Next we construct $G(V, P')$ with this edge weighting function. The resulting co-occurrence graph G' is obtained from $G(V, P')$ by pruning all edges with low weights – here by applying a threshold on a significance measure [3].

II. Higher Order Transformation

A. Definition

We now describe the *higher order transformation* that is used in this work to transform original graphs (of order 1) into higher order graphs (here of order 2 and 3). In the higher order graph, an edge is drawn between two different vertices if they share a common neighbour and attributed a weight according to the number of common neighbours. The adjacency matrix of the higher order graph is obtained by the square of the adjacency matrix of the original graph and placing zeros in the main diagonal (no reflexive edges).

In terms of paths, the higher order transformation is laid out as follows: The above step of constructing G' given G using V and P can be iterated in the following way to construct a sequence $G_0, G_1, \ldots$: While V does not change during the iteration, we need a canonical procedure to define $P = P(G_n)$ at every step. For each $v \in V$, P_v denotes the fully connected graph of neighbouring vertices of v, which are given by $\{a | \{a, v\} \in P(G_n)\}$. Next define $P = \bigcup_{v \in V} P_v$. Using this and the construction above, we start with a set of paths P and define $G_0 = G(V, P)$, then we recursively define G_{n+1} using G_n and its set of local subgraphs $P(G_n)$ as described above. In terms of the adjacency matrix A_n of G_n, we get $A_{n+1} = A_n^2$, but we will apply the following pruning operation.

Since the average path length of scale-free SW graphs is short and local clustering is high, this operation leads to an almost fully connected graph in the limit, which does not allow to draw conclusions about the initial structure. Thus, we prune the graph in every iteration in the following way: For each vertex, only the t outgoing edges with the highest weights are taken into account. Notice that this vertex degree threshold t does not limit the maximum degree, as thresholding is asymmetric. This operation is equivalent with only keeping the t largest entries per row in the adjacency matrix $A = (a_{ij})$, then $A_t = (sign(a_{ij} + a_{ji}))$.

B. Random Graphs of Higher Order

The higher order graph degree distributions of the random graphs of figure 1 show considerable differences. The order 2 BA graph shows a power-law degree distribution with $\gamma = 2$, thresholding has almost no effect. In order 3, the slope is even lower, but the tail is decaying faster than a power-law would predict. The full order 2 DM graph retains its two power-law regimes, yet thresholding leads to a single power-law approximation with $\gamma \approx 2$. In order 3, there is a large quantity of high degree vertices, indicating a densely connected graph.

C. Word Co-occurrence Graphs of Higher Order

Interestingly, the slope of the order 2 BA degree distribution matches the order 1 co-occurrences, cf. Fig. 1 (right). As opposed to this, the degree distribution of order 2 word co-occurrences are influenced by thresholding and exhibit two regions with many low-degree and few high-degree vertices, dependent on t. The full order 2 graph distribution resembles the shape of the order 2 DM graph. But while the order 3 DM graph degree distribution cannot be approximated by a power law, the order 2 co-occurrences show two power-law regimes with $\gamma_1 \approx 1$ and $\gamma_2 \approx 4$ – slopes that are also observed for order 2 DM graphs.

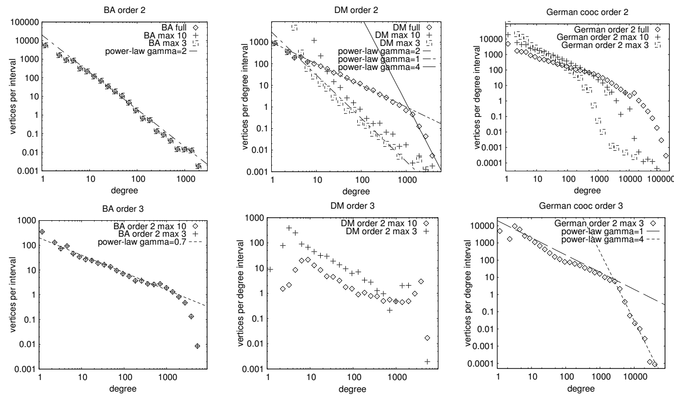

Fig. 2. Orders 2 and 3 for the random graph models and for word co-occurrences for various t

D. Asymptotic Behaviour

Starting with a node v one might be interested in the sequence $P_{n,v}$. For such higher order word co-occurrences we made the following observations in previous experiments:

- As expected, the dynamics often yield to fixed points, i.e. sets of nodes invariant under iteration. Usually, there are several strong attracting fixed points. One can also observe attracting cycles.
- In the case of words, the words in the fixed point are often not semantically related to the starting point, yet the first steps seem more interesting.
- Stronger thresholds lead to fewer fixed points and cycles, the empty set is the only attractor in the limit.

III. Conclusion

We examined the utility of the BA-model and the DM-model to explain the degree distributions of word co-occurrence graphs in natural language. Using a higher order transformation, we could show that both random generators produce graphs that match some of the characteristics of natural language, but no single model can generate graphs – even only with similar degree distributions – that agree with word co-occurrence graphs taking the higher order transformations into account.

References

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

[2] Dorogovtsev, S. N. and Mendes, J. F. F. (2001). Language as an evolving word web. *Proc. Roy. Soc. London. Series B*, 268(1485):2603–2606.

[3] Dunning, T. E. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

[4] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61.

[5] Ferrer-i-Cancho, R. and Solé, R. V. (2001). The small world of human language. *Proc. Roy. Soc. London. Series B*, 268(1482):2261–2265.

[6] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.