Diachronic analysis of co-occurrences networks: a case study on Staribacher diaries and Austrian politics

Matthias Schögl[1], Marcella Tambuscio[1], Maria Mesner[2], and Matthias Trinkaus[3]

[1]Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences
[2]University of Vienna
[3]Kreisky Archives

Josef Staribacher (1921-2014) was an Austrian politician. He was member of the Austrian Social Democratic Party and Minister of Trade, Commerce and Industry in all 4 governments led by Bruno Kreisky. Fortunately he was also an excessive diary writer[1]. He rigorously documented nearly any meeting and discussion he took part in during his 13 years as Austrian minister. His diary contains about 15.000 pages documenting the inner functioning of the Kreisky governments.

We departed from a series of historiographical research questions: In a situation of perceived political crisis (first "oil crisis" 1983/84), who is communicating with whom within the political élite, which political organizations and levels of hierarchy are involved in the decisionmaking process? How did communication patterns change in time (from oil crisis I to oil crisis II? How can these changes be related to a changed political context?

Network science [3] is a simple and powerful tool to represent complex systems of interacting entities. Social Sciences and the Humanities have shown that network science can contribute significantly to understand information flow [13], policy making [9], public opinion making [14] and production of scientific research [7]. Also, there is a huge literature about its usage to represent texts and languages [6, 8, 4, 5]. We focus here on *co-occurrences networks*, where the nodes are words (characters [12] or semantic concepts [2]) and the links represent for the fact that those words appear in the same context. Some previous researches revealed that such networks are quite clustered and simple computations as community detection or centrality measures can be used to extract common features as topic [10] or to unveil patterns that cannot be found with traditional methods based on simple word count statistics [1].

Considering people mentioned in the diaries, we built a co-occurrences network: specifically we use the parse tree of a sentence to find persons who are semantically related and draw an edge between them. We obtained a multi-graph G (more connections among the same couple of nodes) with $N = 4224$ vertices and $M = 68984$ edges. We also incorporated some other information in the graph, labeling edges with two attributes: the date (of appearance in the diaries) and the list of words included among the two occurrences.

Our first goal is to represent and analyze the temporal evolution of these connections. Since the data are related to a quite long time period (a decade), we are particularly interested in exploring the evolution of relationships among the people mentioned in the diaries over time. Then we modeled our network as a sequence of temporal networks, filtering only the most important nodes: after excluding isolated components of the network (people mentioned a few times), we extracted the sub-graphs related to each year and then we considered only the common nodes to all graphs. Finally we obtained a sequence of 11 graphs with 51 vertices and a number of links that varies between 82 and 165. For those 51 persons we labeled the respective node with the list of institutions (political parties, government, department and so on) he or she has been affiliated to.

[1]To be more precise: He did not write the diary himself, but dictated it to his secretary.

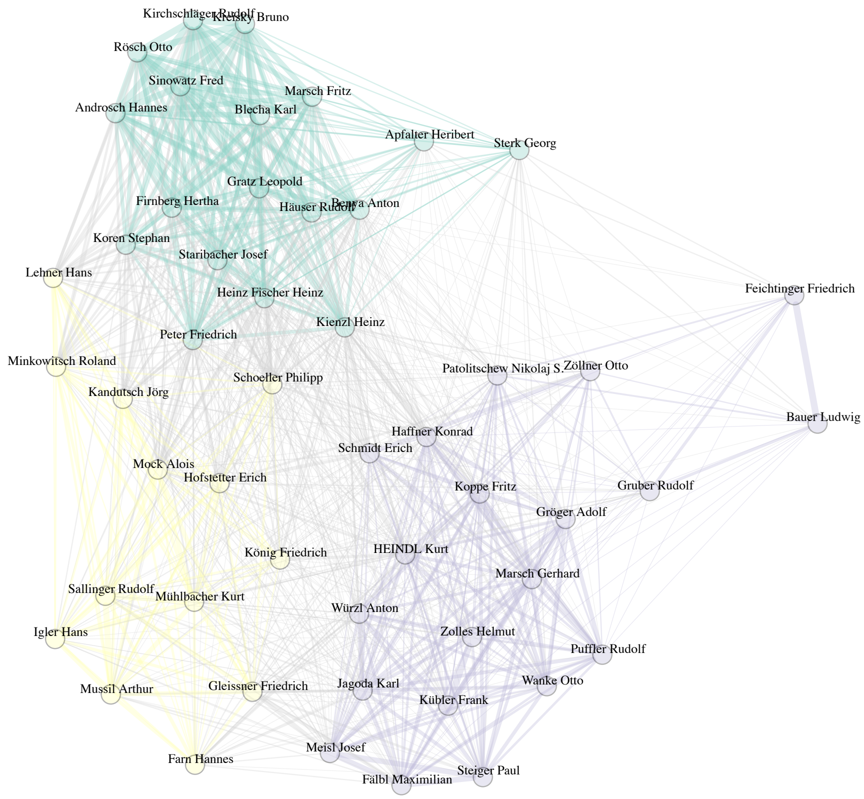

Figure 1: Communities network: nodes are connected if they appear in the same community at least in one temporal network referred to one year in the decade 1971-1981. The color is associated with the results of community detection over these networks: nodes appear to be clustered according to their political alignment.

All these graphs show quite high values of clustering coefficient and exhibit a strong disassortativity [11], meaning that they are clustered. To test this hypothesis we ran community detection algorithm over them and we are currently comparing the results. Interestingly, the communities (more than 6 every year) change over time, then we are interested in exploring how many times two persons appear in the same community, i.e. how strong is the connection among them. To do this, we built another network (see Figure 1 where the nodes are the 51 entities appearing in all the 11 graphs, and each link (i, j) represents the fact that entity i and entity j appear in the same community: the weight on the links w_{ij} keeps tracks of how many times this happens ($w_{ij} \in [0, 11]$). Running again community detection (Louvain algorithm [11] we are able to find 3 groups that reveal the political alignment of involved people. This is actually interesting because the list of people mentioned in the diaries does not include only politicians. Our next research questions include: (i) is there an overlapping among the institutions in the labels and the detected communities?, (ii) can we observe significant modifications in the networks properties that could identify some particular event? Moreover, we are also considering to use the words among the occurrences for further analysis, labeling the relations within topics and exploring their diachronic evolution.

In our presentation/poster we will focus on i) the method to convert the annotated diaries into networks, ii) first experiments with the data and iii) what implications our findings might have for political sciences and how well they align with results from more traditional research projects.

References

[1] D. R. Amancio. Network analysis of named entity co-occurrences in written texts. *EPL (Europhysics Letters)*, 114(5):58005, 2016.

[2] D. R. Amancio, O. N. Oliveira Jr, and L. d. F. Costa. Unveiling the relationship between complex networks metrics and word senses. *EPL (Europhysics Letters)*, 98(1):18002, 2012.

[3] A.-L. Barabási et al. *Network science*. Cambridge university press, 2016.

[4] A. Baronchelli, R. Ferrer-i Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen. Networks in cognitive science. *Trends in cognitive sciences*, 17(7):348–360, 2013.

[5] J. Borge-Holthoefer and A. Arenas. Semantic networks: Structure and dynamics. *Entropy*, 12(5):1264–1302, 2010.

[6] J. Cong and H. Liu. Approaching human language with complex networks. *Physics of life reviews*, 11(4):598–618, 2014.

[7] D. Crane. Transnational networks in basic science. *International Organization*, 1971.

[8] R. F. i Cancho, R. V. Solé, and R. Köhler. Patterns in syntactic dependency networks. *Physical Review E*, 69(5):051915, 2004.

[9] P. Korom. Politisierte Aufsichtsräte? Über die Entwicklung personeller Verflechtungen von Wirtschaft und Politik in Österreich (1989–2011). *Österreichische Zeitschrift für Soziologie*, 39(1):1–22, 2014.

[10] Y. Liu, J. Goncalves, D. Ferreira, B. Xiao, S. Hosio, and V. Kostakos. Chi 1994-2013: mapping two decades of intellectual progress through co-word analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3553–3562, 2014.

[11] M. Newman. *Networks*. Oxford university press, 2018.

[12] B. Rieck and H. Leitte. 'shall i compare thee to a network?': Visualizing the topological structure of shakespeare's plays. 2016.

[13] G. Veltri. Information flows and centrality among elite European newspapers. *European Journal of Communication*, 27(4):354–375, 2012.

[14] D. J. Watts and P. S. Dodds. Influentials, Networks, and Public Opinion Formation. *JOURNAL OF CONSUMER RESEARCH*, 34, 2007.