Towards Visual Sociolinguistic Network Analysis

Kostiantyn Kucher\(^1\), Masoud Fatemi\(^2,3\)\(^b\) and Mikko Laitinen\(^2,4\)\(^c\)

\(^1\)Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden
\(^2\)Center for Data Intensive Sciences and Applications, Linnaeus University, Växjö, Sweden
\(^3\)School of Computing, University of Eastern Finland, Kuopio/Joensuu, Finland
\(^4\)School of Humanities/English, University of Eastern Finland, Kuopio/Joensuu, Finland

Keywords: Social Networks, Social Media, Variationist Sociolinguistics, Social Network Analysis, Network Visualization, Text Visualization, Visual Analytics, Information Visualization.

Abstract: Investigation of social networks formed by individuals in various contexts provides numerous interesting and important challenges for researchers and practitioners in multiple disciplines. Within the field of variationist sociolinguistics, social networks are analyzed in order to reveal the patterns of language variation and change while taking the social, cultural, and geographical aspects into account. In this field, traditional approaches usually focusing on small, manually collected data sets can be complemented with computational methods and large digital data sets extracted from online social network and social media sources. However, increasing data size does not immediately lead to the qualitative improvement in the understanding of such data. In this position paper, we propose to address this issue by a joint effort combining variationist sociolinguistics and computational network analyses with information visualization and visual analytics. In order to lay the foundation for this interdisciplinary collaboration, we analyse the previous relevant work and discuss the challenges related to operationalization, processing, and exploration of such social networks and associated data. As the result, we propose a roadmap towards realization of visual sociolinguistic network analysis.

1 INTRODUCTION

The term social network gained immense popularity during 2000s due to the emergence of Web 2.0 services (Furht, 2010), which allowed users to explicitly denote their relations to other users and explore the relations between other users as well as the digital content created by them (thus leading to social media services). However, research on such relations existing between individuals had already been conducted within sociology for decades by that point (Granovetter, 1973; Scott, 1988; Scott and Carrington, 2011). Analysis of social networks has also provided a useful tool—and corresponding challenges—to researchers in linguistics. More specifically, networks have been studied within variationist sociolinguistics (Milroy and Milroy, 1985; Labov, 2001; Chambers and Schilling, 2013; Laitinen, 2020) as part of the inquiry into the evolution of languages, their use, and variation among individuals, groups, and populations (Milroy, 1980; Milroy, 1992; Marshall, 2004). The main finding in this field is that social networks influence how innovations diffuse into communities. On the one hand, if people are linked with dense and multiplex ties, their networks are close knit, and such structures tend to resist change. On the other hand, network ties can be weak, in which case individuals are predominantly linked through occasional and insignificant ties, and the network is loose knit. Empirical evidence shows that loose-knit networks promote innovation diffusion. This somewhat counterintuitive observation builds on the idea that loose-knit networks consist of people who are on the social fringes, which means that the cost of adopting an innovation is low. Adopting an innovation is socially risky, and people do not want to risk their social standing in close-knit social structures (Granovetter, 1973).

However, the traditional data sets used for social network analysis within variationist sociolinguistics were typically limited to manually collected observations and questionnaires with less than 50–70 individuals (Milroy, 1992; Marshall, 2004). Ample evidence from social anthropology suggests that aver-
age network size, at least in Western societies, tends to be over 150 nodes (McCarty et al., 2001). What is more, the emergence of online social network and social media services has provided the researchers in this community—as well as the social sciences and the humanities in general (Schöch, 2013)—with an opportunity to expand the scope of their analyses to much larger data sets, potentially with additional, rich multivariate information associated with the networks (Hale, 2014; Kim et al., 2014). Such analyses typically rely on computational methods developed within sociology, computer science, and computational sociolinguistics (Jurafsky and Martin, 2009; Newman, 2010; Nguyen et al., 2016), as demonstrated by the recent work (Laitinen et al., 2020).

The enthusiasm for the opportunities promised by big data is sometimes met with a more pragmatic position that stresses the importance of smart data instead (Schöch, 2013). The large scale and the ability to process data sets with high speed are not always sufficient on their own to help the researchers understand the data better, gain useful insights, or formulate further hypotheses, thus highlighting the importance of interactive visual analyses.

In this position paper, we propose to look at the challenge of analyzing social networks for the tasks of variationist sociolinguistics from the point of view of information visualization and visual analytics. By combining the sociolinguistic, computational, and visual perspectives on social networks and the associated multivariate data (which can include geospatial, temporal, and text attributes), we aim to lay the foundation for visual sociolinguistic network analysis (see Figure 1) and raise the awareness of both visualization and sociolinguistic communities.

2 RELATED WORK

In order to understand the challenges and opportunities for our research problem, in this section, we discuss the related work on social network analysis and visualization of network and text data.

2.1 Social Networks in Sociolinguistics

Social networks are studied in variationist sociolinguistics in the context of language use and change (Milroy and Llamas, 2013; Dodsworth and Benton, 2020). By investigating the structure and the patterns of interaction between the individuals and groups in such networks, researchers in sociolinguistics are able to understand how the information propagates, and how it affects the language used by individuals. One of such interaction aspects is related to strong vs weak ties between the members of social networks (Granovetter, 1973), which might lead either to suppression or facilitation of innovative language use in the respective communities. Further aspects taken into account can include latent (long-term and stable) vs emergent (swiftly evolving and renegotiable) networks, coalitions (situational dense clusters), and communities of practice (formed on the basis of certain group activity) (Bergs, 2006), and so on.

In the past decades, the availability of computational methods and digital data from online services has made it possible to investigate not only small-scale networks (including ego networks), but also large simulated networks (Fagyal et al., 2010) and networks extracted from social media data (Grandjean, 2016; Laitinen and Lundberg, 2020; Laitinen et al., 2020). In order to make sense of such larger networks, both computational and visual methods are necessary, which are discussed next.
2.2 Computational and Exploratory Network Analysis

Social networks can be viewed from the more general perspective of network analyses and applied graph theory in computer science (Brandes and Erlebach, 2005; Newman, 2010), with the focus on the topology, structures, and important elements existing in such networks. The last of these tasks can be achieved through the analysis of network centralities for the nodes (such as betweenness or closeness), for instance. In general, multiple network analysis methods have been proposed and applied for social network analyses (Scott, 1988; Aggarwal, 2011; Scott and Carrington, 2011) and relevant aspects of social media and literary data analyses (Agarwal et al., 2012; Pitas, 2016). We should also mention the multilayer network approach (Kivela et al., 2014), which provides a promising unified framework for modeling, analyzing, and representing complex networks.

From the practical point of view, the tools and libraries available for computational analyses of social networks typically include support for multiple tasks, including centrality analysis, community detection, and so on; here, we could list graph-tool (Peixoto, 2014), SNAP (Leskovec and Sosić, 2016), and scikit-network (Bonald et al., 2020) as several examples. Some of the existing tools and libraries also provide at least some capabilities for visualization and exploratory analysis (Brath and Jonker, 2015); here, we could mention the JUNG library (O’Madadhain et al., 2005) and the tools such as Pajek (Batagelj and Mrvar, 2004; de Nooy et al., 2018), EgoNet (McCarty et al., 2007), Gephi (Bastian et al., 2009), NetMiner (Ghim et al., 2014), and NodeXL (Hagberg et al., 2008; Hansen et al., 2011). Some of the approaches developed for network analyses in other domains (e.g., biological network data) have also been successfully applied to social networks, for instance, Zhou et al. describe the application of Cytoscape for social network data analyses as part of the VAST challenge (Zhou et al., 2009). The recent approaches also provide support for computational analyses and visualization of multilayer networks, e.g., MuxViz (De Domenico et al., 2015) and Py3plex (Skrlj et al., 2019). Further contributions on representation and interactive analysis of graphs and networks that originate from the visualization research community are discussed next.

2.3 Network Data Visualization

The recent research efforts in the fields of graph drawing and network visualization cover a number of important tasks for representing and interacting with multivariate networks (Kerren et al., 2014; Nobre et al., 2019), temporal and dynamic graphs (Beck et al., 2014; Kerracher et al., 2015; Beck et al., 2017), group structures (Vehlow et al., 2015), and large-scale graphs (von Landesberger et al., 2011). Several frameworks encompassing multiple aspects of complex real-world networks have been proposed as well, including multi-faceted graph visualization (Hadlak et al., 2015) and multilayer network visualization (McGee et al., 2019).

Some of the tasks even more closely related to social networks in sociolinguistics have also been addressed to some extent by the existing approaches, for instance, visualization of small world networks (Auber et al., 2003; van Ham and van Wijk, 2004) and visual analysis of centralities (Correa et al., 2012; Kerren et al., 2012; Zimmer et al., 2012).

Visualization (Viégas and Donath, 2004) and visual analysis (Zhao and Tung, 2012) of social networks in particular has been in the focus of some previous works, e.g., NodeTrix (Henry et al., 2007) and NodeXL (Bonsignore et al., 2009). Several surveys provide further overview of the approaches existing in this field (Du et al., 2015; Correa, 2017). Finally, we should note that the interest for the challenges of social network visualization exists from the perspective of social media visual analytics (Wu et al., 2016; Chen et al., 2017); network visualization for the humanities (Börner et al., 2019), and visual analysis of multilayer networks across various disciplines and domains (Kivela et al., 2019).

2.4 Text Data Visualization

While the approaches discussed above focus mainly on the network data relevant for sociolinguistics, we should not forget the importance of tasks of visual representation and interaction with language, speech, and text data for this research field. The existing text visualization and visual text analysis techniques have been covered by several existing surveys (Alencar et al., 2012; Kucher and Kerren, 2015; Liu et al., 2019; Alharbi and Laramée, 2019), including the surveys focusing on the more specific problems of topic (Dou and Liu, 2016) or sentiment (Kucher et al., 2018) visualization, or visual analysis of texts for the digital humanities (Jänicke et al., 2015; Jänicke et al., 2017). El-Assady et al. describe their work on a complete software platform that can be used for linguistic analyses (El-Assady et al., 2019); and Hammarström et al. make use of visualization in their work that does not focus on text data per se, but rather the information about the statuses of languages around the world (Hammarström et al., 2018).
As previously mentioned, the information about social networks can also be extracted from rich social media data, which also has been in the focus of multiple visual analytic approaches (Wu et al., 2016; Chen et al., 2017). Several examples of such systems that make use of both network and text data include Whisper (Cao et al., 2012) and Verifi2 (Karduni et al., 2019), among others; and one recent example that focuses on the language use trends on Twitter is Story-wrangler (Alshaabi et al., 2020), for instance.

3 ROADMAP FOR VISUAL SOCIOLINGUISTIC NETWORK ANALYSIS

Based on (1) the prior experiences of researchers in sociolinguistics and computational network analysis and (2) the analysis of the state of the art in InfoVis and visual analytics, we can now propose the roadmap towards realization of visual sociolinguistic network analysis as an interdisciplinary research effort:

Find the Common Ground. In order to establish successful interdisciplinary collaboration, it is important to be aware of the gaps existing between the disciplines and domains (van Wijk, 2006), and to align the goals set by the members of such a collaboration (Kirby and Meyer, 2013). Besides our own prior experiences (Laitinen et al., 2017; Martins et al., 2017; Kucher et al., 2018; Laitinen et al., 2018; Kucher et al., 2020; Simaki et al., 2020), we could also rely on the discussions of previous collaborations between the experts in visualization and the digital humanities (Jänicke, 2016; Hinrichs et al., 2017; Bradley et al., 2018). When designing applications and tools as part of such collaboration, it is also important to consider the gaps between the designers’ and the end-users’ expectations and preferences: in many cases, “simple is good” (Russell, 2016).

Establish the Design Process. The process for discussing the requirements of domain experts and designing solutions proposed by the visualization researchers can be structured according to one of the models proposed in visualization, for instance, Munzner’s nested model for visualization design and evaluation (Munzner, 2009). Since the end goal is not just to design a novel visual representation for network data, but rather to contribute to the efforts by variationist sociolinguistics experts in making sense of complex data from digital sources (Laitinen et al., 2017; Laitinen et al., 2018; Laitinen et al., 2020), the models and workflows discussed in visual analytics must be taken into account, too (Sacha et al., 2014; Andrienko et al., 2018). The visualization design process can also make use of the categorizations of user tasks discussed in literature (Shneiderman, 1996; Brehmer and Munzner, 2013).

Address the Specific Visual Analysis Challenges. Based on the discussion above, we expect that at least the following challenges will have to be tackled in the context of visual analysis of sociolinguistic social networks and the associated data:

- Representation and interaction with multiple (and possibly numerous!) networks, subnetworks, and network elements (Wang Baldonado et al., 2000; Roberts, 2007);
- Comparison of such networks and network elements (Gleicher et al., 2011), including the comparison driven by the centrality analyses and detected group structures (Vehlow et al., 2015);
- Facilitation of the complete visual analytic process for the users (i.e., experts in sociolinguistics), including the tasks of provenance, guidance, and externalization of generated knowledge (Sacha et al., 2014; Andrienko et al., 2018); and
- Integrating the computational and visual analyses of the network data with the corresponding analyses of the associated (meta-)data, which can potentially involve textual (Kucher and Kerren, 2015; Jänicke et al., 2015; Jänicke et al., 2017; Kucher et al., 2018), temporal or dynamic (Cottam et al., 2012; Beck et al., 2014; Kerracher et al., 2015; Beck et al., 2017), as well as geospatial (Dykes et al., 2005) aspects.

This list is not conclusive, of course, and we expect further challenges to be identified in the future.

Evaluate the Resulting Approaches. Besides the challenges of designing and implementing visual analytic solutions for the tasks discussed above, evaluation of such visual analytic solutions is a major challenge on its own (Isenberg et al., 2013). Here, we can use the body of knowledge focusing on task-based user studies and evaluation of visualization and interaction techniques (Purchase, 2012); but also the approaches related to critical discussion (Kosara et al., 2008), reflection (Meyer and Dykes, 2018), expert reviews (Tory and Möller, 2005), questionnaires (Wall et al., 2019), and even crowdsourcing (Archambault et al., 2017) for evaluating InfoVis and visual analytic approaches. Additionally, previous work focusing on design and application studies is also available (Sedlmair et al., 2012; Weber et al., 2017).
Raise Awareness within the Sociolinguistic Community. The last—but definitely not the least important!—part of this roadmap is to use the existing examples, intermediate results, and applications of new approaches to raise awareness about the value and applicability of visualization methods (Fekete et al., 2008) within the sociolinguistic research community. By nourishing such interdisciplinary collaborations (Hinrichs et al., 2017; Bradley et al., 2018), all of the participants can gain new knowledge and progress towards novel, important contributions.

4 CONCLUSIONS

In this position paper, we have discussed the problem of visual analysis of social networks for variationist sociolinguistics. Given the range and the complexity of theoretical and practical challenges for making sense and applying the knowledge about such networks and all of the associated data, including textual, temporal, geospatial, and other aspects, we argue that an interdisciplinary approach is required to tackle this research problem, and that visual analytics should be an integral part of that approach. As our future work, we intend to proceed with realization of the steps listed as our roadmap towards visual sociolinguistic network analysis, and we hope that both sociolinguistic and visualization communities become aware and involved in the work on this interesting and important problem, which can lead to both theoretical findings and practical applications in the future.

ACKNOWLEDGEMENTS

This research has been partially supported by the funding of the Center for Data Intensive Sciences and Applications (DISA) at Linnaeus University. We are extremely grateful for this support.

REFERENCES


Aggarwal, C. C., editor (2011). Social Network Data Analytics. Springer US.


Bergs, A. (2006). Analyzing online communication from a social network point of view: Questions, problems, perspectives. Language@Internet, 3(3).


of the 1st Workshop on Visualization for the Digital Humanities, VIS4DH ’16.
Towards Visual Sociolinguistic Network Analysis