Pie Chart Glyph Visualization of Uncertain Connected Components

Marina Evers[1,2], Farhan Rasheed[3], Talha Bin Masood[3], Ingrid Hotz[3] and Daniel Weiskopf[1]

[1] University of Stuttgart, Germany [2] University of Siegen, Germany [3] Linköping University, Sweden

Abstract

Edges of graphs are often associated with uncertainty. The inherent uncertainty of the data also induces uncertainty in derived graph attributes such as connected components. Even for planar graphs, visualizing the connected components in the graph embedding while encoding their uncertainty imposes challenges due to overlap. We present a visual encoding for uncertain connected components in a planar graph embedding. The underlying model does not require matching or assumptions on the overlap of the components and emphasizes uncertain boundary regions. We discuss different design options and show the applicability of our approach based on synthetic data and real-world data on force networks in granular materials.

1. Introduction

Uncertain graphs occur in different application domains, for example, as force networks in granular material [PMWT05, ZNL*17], where the forces that the individual particles exert on each other can be associated with uncertainty. Extracting connected components of force networks can help identify force chains, which are structures linked to force propagation. However, visualizing the uncertain connected components is challenging if the uncertainty should also be conveyed. Showing small multiples not only imposes challenges on creating a spatial correspondence between the individual visualizations but also does not scale, especially for more complex networks requiring many representations.

Graph uncertainty is often modeled by assigning a probability to the edges [Ban22, KGPT16]. In the context of graph analysis, research focuses mainly on data mining in uncertain graphs, ranging from clustering [CFP*17, HGX*19, YWL*22] over decompositions [BGKV14] to extracting frequent and reliable patterns [JLA11, CZL*19]. Extracting connecting components from uncertain graphs is related to reachability queries and at least as hard to solve as NP-hard problems [JLDW11]. Therefore, the probability distribution of the graph is often represented by samples (also called possible worlds) [PGPB14]. Visualization of uncertain graphs mostly focuses on the uncertainty of edge weights, which are commonly visualized by using visual variables on the edges [vBW17b, CGH*24, GHL15, vKS*11]. Other approaches include the uncertain layout in the visualization by using animation [ZAH22] or splatting and edge bundling [SNG*17]. None of these approaches allows for obtaining information on derived graph properties, such as the location and size of connected components. Encoding the probability of edges neither includes covariances between edges nor shows a clear picture in regions where it is uncertain to which connected component a node belongs.

We consider uncertainty modeled by a probability that indicates for each edge how likely it is to exist. The uncertainty can be conveyed indirectly by providing different possible realizations of the graph or by a probability distribution. In the latter case, we employ Monte Carlo sampling to obtain possible realizations. Then, we extract connected components on each sample individually. We propose a new visual encoding to show these uncertain connected components in planar graphs. We discuss different design options and show the application to uncertain force networks. Note that we consider planar graphs with a known embedding (for example, because the nodes have predefined spatial positions).

2. Visualizing Uncertain Connected Components

To compute the connected components, we work on the samples of the uncertain graph. They can be computed from a probabilistic graph or given, for example, by measurements of an uncertain system. This allows for avoiding increasing complexity by working on the probabilistic graph directly. The final visualization drives our data handling. Ultimately, we want to obtain a visualization that provides an impression of the connected components and their certainty, where small variations should be less prominent compared to larger ones. We considered matching connected components (for example, based on overlap or directly incorporating highly certain edges into the extraction process). However, there is not always an obvious ground truth, which makes it difficult to establish robust matching heuristics. Instead, we independently compute the connected components for each sample of the uncertain graph and use visual similarities to build correspondences (see Figure 1a).

We choose a glyph visualization as it allows for combining the spatial encoding of the node positions while additionally showing the connected components [BKC*13]. We draw a pie chart for each node to visualize the connected components. A pie chart glyph was previously used to encode fuzzy community memberships [VRW13]. In our case, the portion of the glyph that is colored

DOI: 10.2312/evp.20251131

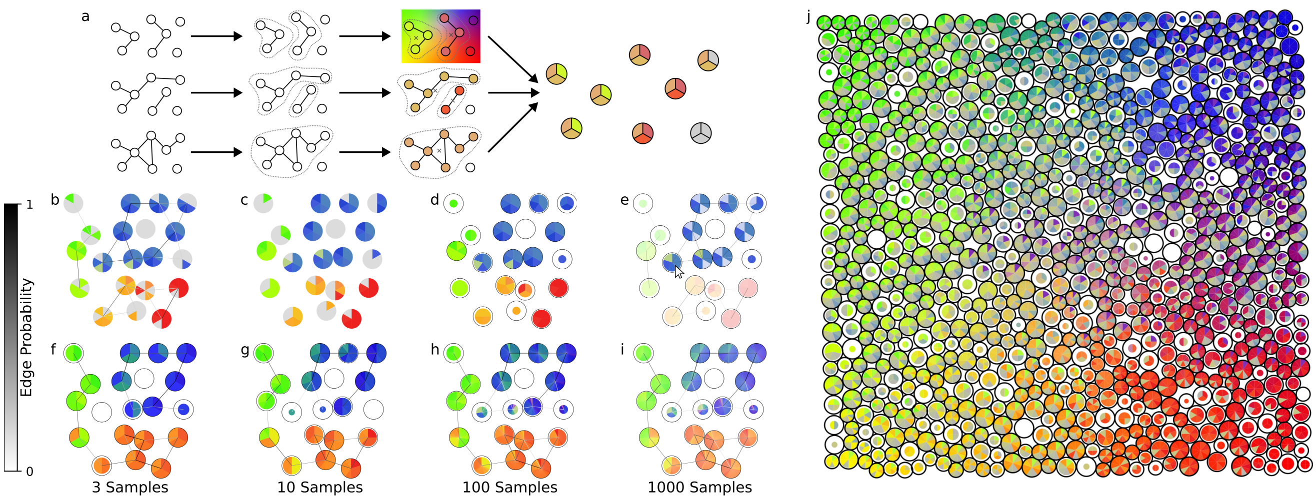

Figure 1: *The schematic figure (a) shows how the visualization is created. Different design options for uncertain connected components: (b) Colors within the disk-shaped glyphs indicate to which component the nodes belong. The darkness of the edges indicates probability. (c) The slices can be ordered such that similar colors are grouped together. (d) Sizes can be used to encode the probability of the nodes being connected to other nodes. (e) Hovering over a node allows for investigating the corresponding connected components in more detail. (f–i) 3 to 1000 samples of the same synthetic dataset indicate a good scalability with the number of samples. (j) Application example of connected components in a force network in granular material.*

in the same color represents the portion of samples that belong to the respective connected component. Our goal is to assign similar colors to connected components with large overlaps and different colors to connected components that are spatially separate. We achieve this goal by using the barycenter of all nodes belonging to the connected component to assign a color based on a 2D colormap [SBT*15, RNN*24]. If a node is not connected to any other, the corresponding slice is shown in light gray to visually encode that the node does not contribute to the network structure.

Optionally, the edges can be shown as an overlay to the pie charts. Here, the probability that the corresponding edge exists is shown using a grayscale colormap. Figure 1b shows an example of this visual encoding. Ordering the slices in the pie chart according to the samples also implicitly encodes correlations. However, it might be more difficult to interpret percentages. Therefore, we also provide the option to order the individual slices based on the similarity of the colors starting in the top-center. The slices indicating non-connected nodes are placed at the end (see Figure 1c). To highlight how likely it is for a particle to belong to a connected component, size can be used to encode this probability. As sizes of nodes might be used to encode other information, such as radii of particles in granular material, we show the outer radius of the pie chart by using a black circle (see Figure 1d). Users can interactively switch between the different representations. For a more detailed analysis, it is also possible to hover over a node, and slices of pie charts that belong to the same connected component are highlighted by decreasing the saturation of the other elements (see Figure 1e).

3. Results

One criterion for designing the visualization was that it scales well if the number of samples of the uncertain graph is increased. We test the scalability [RPA*24] by visually comparing the results for varying numbers of samples from synthetic data. The results are shown in Figures 1f–i. Even for 1000 samples, different structures are clearly visible. Figures 1h and i indicate convergence for large sample numbers, as there is no substantial difference between the percentages in similar colors. We also apply our visualization to measured data of 2D granular material [RMM*23, RNN*24], where 10 measurements of the same system consisting of 832 disks (nodes in the force network) are taken. We use the 10 graphs obtained from the measurements as samples representing the uncertain graph. The results are shown in Figure 1j. Using the size encoding, it is clearly visible which disks do not participate in the force network. The colors also reveal that several more locally connected components dominate the uncertain network. However, some of the measurements show larger connected components spanning most of the domain.

In future work, we plan to optimize the visualization by considering perceptual differences in the underlying color map and provide more advanced filtering options. We also plan to include the visualization in an interactive system for a more comprehensive visual analysis of uncertain graphs and evaluate our visual design together with domain experts working with this kind of data. Also, other types of visual encodings of group structures within graphs [VBW17a] could be investigated for their suitability for uncertain graph components.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

References

[Ban22] BANERJEE S.: A survey on mining and analysis of uncertain graphs. *Knowledge and Information Systems 64*, 7 (2022), 1653–1689. doi:10.1007/s10115-022-01681-w. 1

[BGKV14] BONCHI F., GULLO F., KALTENBRUNNER A., VOLKOVICH Y.: Core decomposition of uncertain graphs. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), KDD '14, Association for Computing Machinery, pp. 1316–1325. doi:10.1145/2623330.2623655. 1

[BKC*13] BORGO R., KEHRER J., CHUNG D. H. S., MAGUIRE E., LARAMEE R. S., HAUSER H., WARD M., CHEN M.: Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports* (2013), Sbert M., Szirmay-Kalos L., (Eds.), The Eurographics Association. doi:/10.2312/conf/EG2013/stars/039-063. 1

[CFP*17] CECCARELLO M., FANTOZZI C., PIETRACAPRINA A., PUCCI G., VANDIN F.: Clustering uncertain graphs. *Proceedings of the VLDB Endowment 11*, 4 (2017), 472–484. doi:10.1145/3186728.3164143. 1

[CGH*24] CONROY M., GILLMANN C., HARVEY F., MCHEDLIDZE T., FABRIKANT S. I., WINDHAGER F., SCHEUERMANN G., TANGHERLINI T. R., WARREN C. N., WEINGART S. B., REHBEIN M., BÖRNER K., ELO K., JÄNICKE S., KERREN A., NÖLLENBURG M., DWYER T., EIDE Ø., KOBOUROV S., BETZ G.: Uncertainty in humanities network visualization. *Frontiers in Communication 8* (2024). doi:10.3389/fcomm.2023.1305137. 1

[CZL*19] CHEN Y., ZHAO X., LIN X., WANG Y., GUO D.: Efficient Mining of Frequent Patterns on Uncertain Graphs. *IEEE Transactions on Knowledge and Data Engineering 31*, 2 (2019), 287–300. doi:10.1109/TKDE.2018.2830336. 1

[GHL15] GUO H., HUANG J., LAIDLAW D. H.: Representing Uncertainty in Graph Edges: An Evaluation of Paired Visual Variables. *IEEE Transactions on Visualization and Computer Graphics 21*, 10 (2015), 1173–1186. doi:10.1109/TVCG.2015.2424872. 1

[HGX*19] HAN K., GUI F., XIAO X., TANG J., HE Y., CAO Z., HUANG H.: Efficient and effective algorithms for clustering uncertain graphs. *Proceedings of the VLDB Endowment 12*, 6 (2019), 667–680. doi:10.14778/3311880.3311884. 1

[JLA11] JIN R., LIU L., AGGARWAL C. C.: Discovering highly reliable subgraphs in uncertain graphs. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2011), KDD '11, Association for Computing Machinery, pp. 992–1000. doi:10.1145/2020408.2020569. 1

[JLDW11] JIN R., LIU L., DING B., WANG H.: Distance-constraint reachability computation in uncertain graphs. *Proc. VLDB Endow. 4*, 9 (2011), 551–562. doi:10.14778/2002938.2002941. 1

[KGPT16] KASSIANO V., GOUNARIS A., PAPADOPOULOS A. N., TSICHLAS K.: Mining uncertain graphs: An overview. In *International Workshop of Algorithmic Aspects of Cloud Computing* (2016), Springer, pp. 87–116. doi:10.1007/978-3-319-57045-7_6. 1

[PGPB14] PARCHAS P., GULLO F., PAPADIAS D., BONCHI F.: The pursuit of a good possible world: Extracting representative instances of uncertain graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 2014), SIGMOD '14, Association for Computing Machinery, pp. 967–978. doi:10.1145/2588555.2593668. 1

[PMWT05] PETERS J., MUTHUSWAMY M., WIBOWO J., TORDESILLAS A.: Characterization of force chains in granular material. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 72*, 4 (2005), 041307. doi:10.1103/PhysRevE.72.041307. 1

[RMM*23] RASHEED F., MASOOD T. B., MURTHY T. G., NATARAJAN V., HOTZ I.: Multi-scale visual analysis of cycle characteristics in spatially-embedded graphs. *Visual Informatics 7*, 3 (2023), 49–58. doi:10.1016/j.visinf.2023.06.005. 2

[RNN*24] RASHEED F., NASEER A., NILSSON E., MASOOD T. B., HOTZ I.: Multi-scale cycle tracking in dynamic planar graphs. In *2024 IEEE Topological Data Analysis and Visualization (TopoInVis)* (2024), IEEE, pp. 44–54. doi:10.1109/TopoInVis64104.2024.00009. 2

[RPA*24] RICHER G., PISTER A., ABDELAAL M., FEKETE J., SEDLMAIR M., WEISKOPF D.: Scalability in visualization. *IEEE Transactions on Visualization and Computer Graphics 30*, 7 (2024), 3314–3330. doi:10.1109/TVCG.2022.3231230. 2

[SBT*15] STEIGER M., BERNARD J., THUM S., MITTELSTÄDT S., HUTTER M., KEIM D. A., KOHLHAMMER J.: Explorative analysis of 2D color maps. In *WSCG 2015 Conference on Computer Graphics, Visualization and Computer Vision* (2015), pp. 151–160. 2

[SNG*17] SCHULZ C., NOCAJ A., GOERTLER J., DEUSSEN O., BRANDES U., WEISKOPF D.: Probabilistic Graph Layout for Uncertain Network Visualization. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 531–540. doi:10.1109/TVCG.2016.2598919. 1

[VBW17a] VEHLOW C., BECK F., WEISKOPF D.: Visualizing group structures in graphs: A survey. *Computer Graphics Forum 36*, 6 (2017), 201–225. doi:10.1111/CGF.12872. 2

[vBW17b] VON LANDESBERGER T., BREMM S., WUNDERLICH M.: Typology of Uncertainty in Static Geolocated Graphs for Visualization. *IEEE Computer Graphics and Applications 37*, 5 (2017), 18–27. doi:10.1109/MCG.2017.3621220. 1

[vKS*11] VON LANDESBERGER T., KUIJPER A., SCHRECK T., KOHLHAMMER J., VAN WIJK J., FEKETE J.-D., FELLNER D.: Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Computer Graphics Forum 30*, 6 (2011), 1719–1749. doi:10.1111/j.1467-8659.2011.01898.x. 1

[VRW13] VEHLOW C., REINHARDT T., WEISKOPF D.: Visualizing fuzzy overlapping communities in networks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2486–2495. doi:10.1109/TVCG.2013.232. 1

[YWL*22] YU D., WANG D., LUO Q., ZHENG Y., WANG G., CAI Z.: Stable structural clustering in uncertain graphs. *Information Sciences 586* (2022), 596–610. doi:10.1016/j.ins.2021.11.078. 1

[ZAH22] ZHANG D., ADAR E., HULLMAN J.: Visualizing Uncertainty in Probabilistic Graphs with Network Hypothetical Outcome Plots (NetHOPs). *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2022), 443–453. doi:10.1109/TVCG.2021.3114679. 1

[ZNL*17] ZHANG L., NGUYEN N. G. H., LAMBERT S., NICOT F., PRUNIER F., DJERAN-MAIGRE I.: The role of force chains in granular materials: from statics to dynamics. *European Journal of Environmental and Civil Engineering 21*, 7-8 (2017), 874–895. doi:10.1080/19648189.2016.1194332. 1