Network Patterns and Style Contagion: Tracing Artistic Links through Time and Style

Asset Kabdula kabdula_asset@student.ceu.edu
Mihaly Hanics hanics_mihaly@student.ceu.edu
Nurbek Bektursyn bektursyn_nurbek@student.ceu.edu

Introduction:

This research project examines the network of painters starting from the 13th to the 20th century. It mainly focuses on understanding the artistic influence and collaboration at the time, as mentioned earlier, via analyzing open/closed triangles, homogeneous/heterogeneous connections, and creating a framework for linking models (such as Erdős-Rényi random connections, preferential attachment, triangular closure) to see which model is the best for recreating the network. Combining painting datasets to gather painter data lets us collect and work with a broad range of attributes, enabling us to do more exhaustive research to understand the factors that have impacted the painting world throughout history. Learning the interconnectedness and impact of artists on one another can allow us to gain insights into the development of communities (long-lasting communities over centuries) and underlying transmission processes such as artistic social contagion. Our motivation was to know more about artistic movements and gather insight into dynamic networks that developed over a much extended period than an active agent in the network. This project can be attractive to social scientists, art researchers, and theoretical modelers of networks, as research on modeling such networks is not very broad.

The dataset and network:

As there were no painter networks of considerable size and painter datasets were also limited or not public, we combined data from paintings from three different sources: WikiArt, Art500k, and Wikipedia. The WikiArt dataset contains very well-maintained data, with good information on the styles and movements of each picture. However, the Art500k dataset includes much more diverse information on pictures, often data about the author. The problem is that it could be sparser and contain faulty data and other minor issues. We altered through paintings, collecting information about the pictures' painters and from the artists' cross-section in both datasets. By collecting attributes from both datasets and through our own defined attributes, we created wealthy data of 2457 painters. This contains most painters and is reliable, with much fewer errors than the Art500k dataset (primarily due to the painters appearing in both datasets having a big enough fame and well-kept data). We added extra information from Wikipedia data fetching, such as birthplace (later processed with SpaCy NLP package) and some manually added data on a few "corrected" painters. We thus have 21 attributes: nationality, birth, places (with years), styles, movement, coworkers, influences, and so on.
For this project, we created a dynamic network of painters: we connected painters roughly if they painted at the same place at the same time, and we made this network dynamic by adding a time component, adding painters to the network subsequently in the order that they were born. These two attributes (location and time) were considered, and style not, because when we analyze stylistic connections, the network would be biased (higher percentage of links between same style/movement painters). All code (+data, networks) are on GitHub.

Analyses and results:

The existing network was initially analyzed through the number of open and closed triangles (see Figure 1). During earlier centuries, the number of closed triangles stayed relatively steady, and their presence was more dominant than that of open triangles, illustrating the common types of artistic collaborations, such as guilds and workshops, that were popular then. Such collaborations were like a mentoring system, where new painters were directly affected by old experts, fostering direct and robust ties. However, open triangles did happen and indicated the presence of situations in which an artist was influenced or cooperated with two people who did not have a direct link. For example, this may happen if an artist trained with a single instructor before relocating to another location and working with a different one, bridging two unrelated nodes in the network. The sharp rise in closed triangles in the late nineteenth and early twentieth centuries, on the other hand, was likely related to the creation of various art movements and schools, where a variety of artists had the opportunity to move freely between cities and exchange ideas with a broader array of fellow students and teachers.

To shed light on the structural differences between a real-world artistic network and an artificial random network, we found it necessary to create a random graph employing the Erds-Rényi model, with the edge probability p set to around 0.02, to ensure a comparable density of connections as the initial graph. The random graph had a substantially smaller number of closed triangles than the original one, indicating an absence of close-knit clusters common in real-world scenarios. This might be because the random graph does not have historical, social, and cultural constraints that force artists to create solid and collaborative bonds in the original network. Thus, the analysis confirmed that our original network has non-random and complicated properties. It emphasizes the significance of direct and indirect links between artists and how these ties have affected the development of art history.

The following analysis is an extension of the previous analysis but from the perspective of social contagion. According to the social contagion theory, ideas and techniques can spread like a community virus. In this case, the different ideas and methods of artists could be spread and shared among connected artists, which may answer the question of whether the spreading of ideas within artists' community leads to the formation of particular dominant movements or it leads to the appearance of different independent and diverse movements. Homogeneous connections were defined as artists belonging to the same movement, originating from the same time period and location, while heterogeneous connections were those with different movements. This analysis explores the concept of homogeneous connections within art movements, highlighting the potential for collaboration and influence among artists from the same stylistic or thematic background.

The analysis investigated the prevalence of homogeneous connections and their implications for the evolution of artistic movements. The data indicates a dynamic artistic landscape, with many homogeneous connections suggesting a rich tapestry of collaboration and shared influences. The graph of homogeneous and heterogeneous connections shows a steep increase in both homogeneous and heterogeneous connections. In contrast, the graph with the percentage of homogeneous connections depicts a declining trend in the percentage of homogeneous connections, reflecting increased diversity. The findings suggest that while there's a substantial foundation of homogeneous connections, the artistic scene is increasingly diversified, indicating a breakdown of traditional movements and a rise in hybridization and cross-movement influences.

To find what types of connections painter may have in this network (analyzing on a macro-scale), we need to create models that only let nodes make connections in varying, short time intervals. Think of it as if we would only build a "sliding window of the complete network" at a time. For this, we created the TT-framework (Turnoff & Time interval), which takes some linking algorithm (preferential attachment) and "intervals" (of data), the number of nodes and edges, interval endpoints, and the average lifetime of nodes given for each interval. Basically, in each time step (in our case, for every year), the algorithm checks which interval we are currently in inside this step, based on the average nodes "born" in a time step (nodes divided by length of interval) randomly decides how many nodes are born, if any, uses the linking algorithm to connect (or not connect) it to "active" nodes, and then randomly decide for each node if it is turned off (using the average lifetime), then go to the next time step. We used this framework to recreate our graph: we divided the network into three disconnected components (renaissance, baroque, and everything after them), then the networks into 3-2-7 intervals, respectively, based on observed hubs and artistic periods, and computed the inputs to our framework. We implemented three linking methods: Erdős-Rényi (random) linking, preferential attachment, and preferential attachment with one triangle closure and tested them in the framework. We ran these models on all three networks.

We evaluated the resulting graphs in average clustering coefficient and graph edit distance from the original graph. From the results, we have found some things common for all three networks and some things specific to a network. Even though the random linking model is faulty implemented as it tends to have 10-20% more edges than it should have (we have two implementations, one derived from theory with probabilities, one is a simple estimation of in a step how many edges should be constructed, both models overestimate, probably the interconnection of intervals is too dense), it did not seem to be a suitable model in any case, the clustering coefficient is always multiple times lower. Preferential attachment barely had higher clustering coefficients in all cases but always had a "closer" graph to the original (at least by distance). Interestingly, the last, triangle closure implementing model (which adds one less preferentially attached edge and closes a triangle instead) has a much closer clustering coefficient than other models and creates the closest (only 2% better than basic preferential attachment) network to the renaissance network, it is the poorest model for the baroque network by distance. As we look into the network, we see that whilst there is a high amount of triangular closure, since the network is so dense (even random linking had a high clustering coefficient), other length closures are even more common, and thus, closing triangles isn't optimal here. Simple preferential attachment led to a 17% better model. Whilst we did not allocate enough computational power, time to have an estimate of how well the "last" graph models did, with the decrease of triangular closure comes a decrease in this model's accuracy too.

Evaluation and future work:

We have found that different periods had different types of structures, with later periods being less reliant on triangle closure and same-style connections between artists. For future work, expanding the dataset to more painters than the cross-section of the datasets would provide broader, and analyzing additional attributes would provide deeper insights. Exploring more complex linking models that incorporate attributes could also refine the understanding of the network evolution, and could further light up the complex web of historical and social factors that have shaped the world of painting.

Appendix:

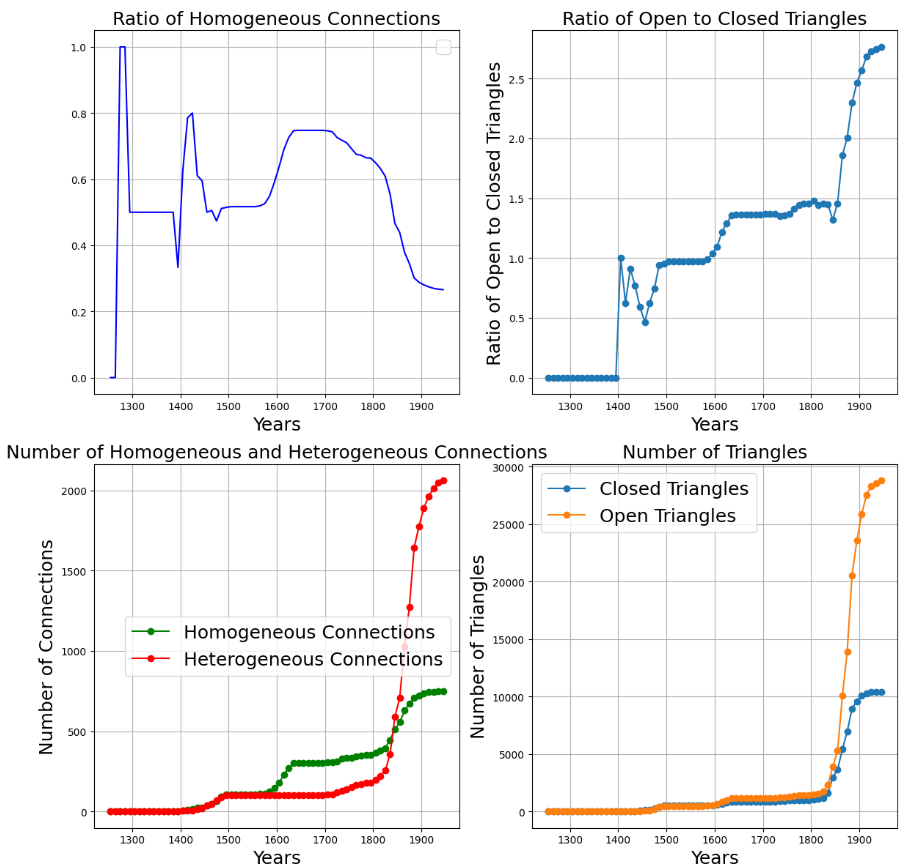

Figure 1: The ratio and number of connections and triangles. There is a striking similarity between the curve of open triangles and the curve of heterogeneous connections. Note: Every artist and his connections are added to the network in the year he was born. This may be misleading, as they were not active at birth, but only later, and one may have to look at time values with a 20+ year delay.

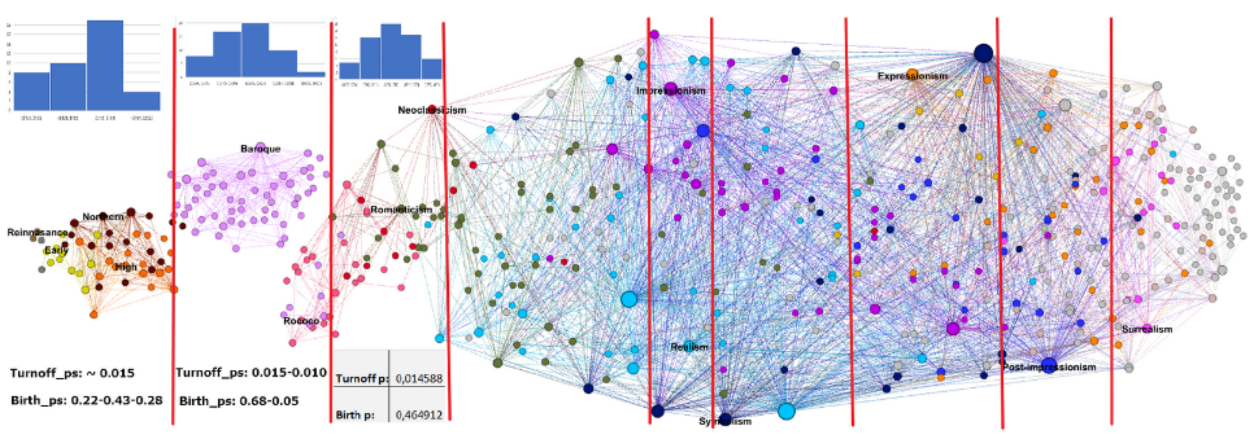

Figure 2: The network divided into time intervals, with example turnoff and birth probabilities. Note: Every artist and his connections are added to the network in the year he was born.