Family Resemblance in Genre Stylistics
A Case Study with Nineteenth-Century Spanish-American Novels

Ulrike Henny-Krahmer

Abstract Family resemblance is a concept that has been introduced into genre theory as an analogy to describe partial and overlapping similarities between different works of the same genre. The concept aims to capture continuities and shifts in historically changing realizations of genres. In this article, family resemblance is applied in a digital genre stylistics analysis of subgenres of nineteenth-century Spanish-American historical and sentimental novels. A formal implementation of the concept is developed, and its usefulness in a corpus-based and quantitative setup is tested. For this, networks of nearest neighbors are built on topic features, and subgroups in the networks are identified with community detection. As a result, the concept of family resemblance itself undergoes change, but the digital methods also appear in a new light as tools that can be engaged for soft categorization.

Keywords genre stylistics, genre theory, categorization, family resemblance, nineteenth century, Spanish-American novel, subgenres, network analysis, topic modeling

1. Genre Categories, Family Resemblance, and Digital Genre Stylistics

A central aspect of literary genre theory is the discussion about the ways in which genres can be defined. More precisely, this includes the question of what kind of categories genres can be understood and conceived as. In this context, it is assumed that the function of generic terms is to represent a group of texts associated with them adequately. The question about the categorial status of genres is of importance both

Henny-Krahmer, Ulrike: Family Resemblance in Genre Stylistics. A Case Study with Nineteenth-Century Spanish-American Novels, in: Hesselbach, Robert, et al. (Eds.): Digital Stylistics in Romance Studies and Beyond, Heidelberg: Heidelberg University Publishing, 2024, pp. 149–171. DOI: https://doi.org/10.17885/heiup.1157.c19370

on a systematic level as well as from a historical perspective. The system is paramount, for example, when literary scholars define generic terms to capture recurrent characteristics inside groups of texts and the differences between them. On the other hand, the historical perspective prevails when scholars analyze and reproduce how historical genre labels related to groups of texts that they were associated with by writers, readers, and critics of the time.[1] Dependent on the dominant perspective, different concepts of genres as categories prevail. Two basic ways to categorize can be distinguished: classificatory and typological categorization. In their purest form, classifications lead to disjunct groups of texts and do not allow for overlaps. Types, on the other hand, form the basis for fuzzy categories because a text can be more or less similar to an ideal type at the center of a category. Both ways to categorize have been advocated for and applied in genre theory and history (Müller 2010, 21; Strube 1993, 59–65; Tophinke 1997). Furthermore, the tension between strict classificatory terms and historical terms that are collectively coined, anchored in time, and hence unsharp has led to the development of flexible approaches to genre definitions. These combine necessary and optional features that texts need to have in order to be considered instances of a genre (Fricke 2010, 7–10).

An approach that completely abandons the idea of necessary common features is the semantic concept of family resemblance, which the philosopher of language Ludwig Wittgenstein developed:

> Consider, for example, the activities that we call “games”. I mean board-games, card-games, ball-games, athletic games, and so on. What is common to them all?—Don’t say: “They must have something in common, or they would not be called ‘games’”—but look and see whether there is anything common to all. For if you look at them, you won’t see something that is common to all, but similarities, affinities, and a whole series of them at that […]. And the upshot of these considerations is: we see a complicated network of similarities overlapping and criss-crossing […]. I can think of no better expression to characterize these similarities than “family resemblance”; for the various resemblances between members of a family—build, features, colour of eyes, gait, temperament, and so on and so forth – overlap and criss-cross in

[1] Ultimately, these two perspectives cannot be strictly separated from each other because literary scholars setting up definitions of literary genres will not completely ignore historical conventions of the terms, as these are at least in part also transmitted through literary history. On the other hand, historical labels are not entirely free of any systematic relationship to textual patterns, either. At the same time, both complexes are confronted with historical changes affecting the relationship between the terms and the subjects. However, there has been much discussion about the distinction and mediation of genre theory and genre history, see Neumann and Nünning (2007, 9–10).

> the same way.—And I shall say: “games” form a family (Wittgenstein 2009, §66–67).

In Wittgenstein’s work, the concept serves as an analogy to describe linguistic activities involving the use of the same word for phenomena with partial and indirect similarities, like the ones seen between different family members. This analogy was adopted in literary genre theory in the 1960s and it became popular because it allowed for more open definitions of genre. According to the family resemblance concept, not all genre-relevant features need to be present in all literary works assigned to it. However, the concept was also criticized as being too loose, mainly because the boundaries between different categories are not defined sharply (Fishelov 1993, 53–68; Fricke 2010, 8–9).[2] Nonetheless, its potential value for genre theory has been emphasized: “I believe that genre theory within literary studies can, on the basis of the concepts of family resemblance and prototypes, manage to realign key questions, especially those arising from the polysemy and historicity of genre concepts” (Hempfer 2014, 414).[3]

The categorization of literary texts by genre is also one of the key concerns of digital stylistics but the discussion about the categorical status of genres that took place in literary genre theory is usually not directly addressed in digital approaches. In many literary stylistic papers, classificatory approaches are used, focusing on which features are most suitable to capture differences between genres. As features, such studies include aspects of style and content, but also structural characteristics of the texts such as text order or representation of speech (Calvo Tello 2018; Gianitsos et al. 2019; Henny-Krahmer 2018; Hettinger et al. 2016; Schöch 2017; Schöch et al. 2016; Underwood 2015). Other studies concentrate on differences between authors, not genres (Calvo Tello et al. 2017; Schöch 2013). Digital studies on the classification of genres, in the strict sense of logical classes, are relevant for a number of reasons. They help to model and interpret textual cues that are crucial to recognizing genres. They also contribute to assessing established methods of text mining, machine learning, and natural language processing (NLP) regarding their value for genre classification. When tested empirically on corpora of different languages, periods, cultural contexts, and genres,

[2] Hempfer (2014, 409–10) points out that, on the other hand, it has been misinterpreted by several genre theorists trying to lead it back to require overall fundamental traits for genres.

[3] As an example of the application of the family resemblance concept, Hempfer (2014, 416–17) describes the history of the elegy, a genre that was originally only identifiable metrically and later by several other traits, amongst other things, intertextual references and motifs. He concludes: “The diachrony of the genre can best be represented as a synchronic network of relations, in which each individual text or epochal version of the genre is linked to other historical versions through common features. […] The genre identity, then, is not produced by a single trait but by the entirety of all relations among their historical versions” (2014, 419). For an application of the family resemblance concept to genre theory, see also Strube (1993, 21–25), who interprets a definition of the novella set up by Seidler in that way.

they expand knowledge about the extent to which the methods are sensitive to the kind of data. In addition, they also help to solve practical problems such as, for example, indexing large collections of texts.

Even so, it is important to question the composition of the categories examined. If the results are low classification rates, for instance, they cannot only indicate problems with the selection of features or classification parameters and methods, but may also be due to the underlying category not having the structure that is assumed. Alternative categorization methods can therefore be a good complement to the *classic* classification methods. It is not that only traditional classification methods have been used in digital genre stylistics. Thoughts about prototypicality, historical variability, and social embedding of literary genres have been expressed and tested as well. Underwood (2016; 2019, 34–67), for example, approaches genre via the history of reception, comparing competing definitions of detective fiction, science fiction, and the Gothic with statistical methods. Henny-Krahmer et al. (2018) use measures of similarity to determine the distances of novels belonging to different subgenres to predefined prototypes. In his contribution to this volume, Schröter (2023) proposes applying machine learning methods to reconstruct the historical change of *disordered* genres such as the German *Novelle*. Such experiments link key discussions of literary genre theory to digital genre stylistics. They contribute to a deeper confrontation of research results in the areas involved, potentially increasing their interest in each other and also challenging the findings achieved in one of the disciplines. For example: are the computational methods really suited for the analysis of genre concepts beyond classification, which is the prevalent method for categorization in computer science? The mentioned papers indicate that they are, but that the methods need to be adapted or creatively used to that end. On the other hand, what happens to the literary theoretical concepts of genre if they are tested for applicability in empirical digital studies? They may need to be modified, and furthermore, genre theory can receive new input from the results of digital genre stylistics. Of course, there is a similar relationship of exchange between literary genre theory and historical and empirical research in literary studies, but the difference to digital stylistics lies in the methodological apparatus. In the latter case, it is highly influenced by computational linguistics and computer science, disciplines with different scientific-theoretical backgrounds, which makes this kind of exchange equally promising and challenging.

This article aims to contribute to the described interface between literary genre theory, literary historical research, and digital genre stylistics. To that end, the concept of family resemblance is formalized and applied in a case study on subgenres of nineteenth-century Spanish-American novels. The choice of the corpus is due to the disciplinary orientation of this article's author and is explained further in the next section. The approach chosen to map the idea of family resemblance to a digital text analysis environment is network analysis. This way of formalization appears highly suitable, given the explanation of family resemblance formulated by Wittgenstein.

2. Subgenres of Spanish-American Nineteenth-Century Novels

Novels lend themselves well to an analysis based on the concept of family resemblance because the novel as a genre can hardly be defined uniquely and in formal terms. The necessary formal conditions are not sufficient to distinguish novels from other fictional narrative prose texts of considerable length (Hempfer 2014, 410), nor can additional formal or content-related criteria serve to capture all types of novels (Fludernik 2009, 627). As Fowler (1982, 112–13) points out, subgenres, while usually sharing the formal features of the genre, are often determined by subject matter or motifs and can be specified from a whole range of perspectives and on increasing levels of detail. He continues:

> Subgenres also threaten to defy subdivision in that they are extremely volatile. To determine the features of a subgenre is to trace a diachronic process of imitation, variation, innovation—in fact, to verge on source study. At the level of subgenre, innovation is life. Here, simple resemblance hardly produces a new work: at the very least there is elegant variation. And from time to time quite fresh subgenres will be invented, enlarging the kind in new directions altogether. It may be the conventionality of subgenres that strikes the beginner. But in reality they are the common means of renewal (Fowler 1982, 114).

As not even the genre itself is unified, the degree of variation applies even more to the subgenres of the novel (ibid, 118–126).

Early Spanish-American novels were influenced by European models, including types of the romantic novel such as sentimental novels and historical novels.[4] Often, the subgenres were adjusted to better serve the needs of expression of the Spanish-American authors. For example, in the nineteenth-century novels, political issues were often mixed up with a love story. The most prominent example is the novel *Amalia* (1851–1855) by the Argentine José Mármol, which tells the story of a group of resistance fighters against the dictatorship of Rosas[5] and of the protagonist's tragic love relationship (Dill 1999, 127). This novel is also an example of a special type of historical novel practiced in nineteenth-century Spanish-America because the contemporary

[4] Nineteenth-century novels are called *early* here because, in the Spanish-American colonies and countries, the genre only took off considerably in the course of that century (see Gálvez 1990, 15–25; Lindstrom 2004).

[5] Juan Manuel de Rosas (1793–1877) was a governor of the province of Buenos Aires who established a dictatorial system marked by repressive measures that lasted between 1829 and 1852 and that enforced a political and economic hegemony of Buenos Aires over the other provinces.

political events are presented as if they were historical. These *prospectively historical* novels (Molina 2011, 285–312) helped the authors to conceal that they were actually criticizing current political regimes and circumstances, and they served to inscribe events of the present or recent past into the history of the country. Then again, there were also conventional types of sentimental and historical novels, the latter, for example, dealing with the history of America's conquest, colonial times, and also European history (see, for instance, Read 1939).

Although more types of subgenres of the novel were practiced by Spanish-American writers in the nineteenth century and interpreted by literary historians later on, this article focuses on novels that have been primarily described as having a sentimental or historical theme. These descriptions may have been made either explicitly or implicitly, by contemporary authors or by modern critics. No distinction will be made here between different historical perspectives on the works, mainly because there is not much dissent for these subgenres and not enough data. Not all the novels were categorized by their authors, and especially lesser-known works have not been discussed extensively by literary historians. Novels from three countries, Argentina, Mexico, and Cuba, were chosen to cover different geographical and sociocultural areas of Spanish-America. The novels selected were first published between 1840 and 1910.[6]

The first aim of the family resemblance analysis is to find out how the two subgenres are organized internally: by looking at the network of similarities between the individual novels of a subgenre, do subgroups, i.e., *families*, emerge? Which traits hold them together? Can prospectively historical novels, for example, be distinguished from other types of historical novels? Or are there differences by country or over time? No preliminary assumption is made here as to the kind of connections that the family resemblance analysis might reveal. There could be diachronic shifts but also synchronic variations in the subgenres. In a second step, the sentimental and historical novels are analyzed together to see how the two subgenres are connected when no strict boundaries are applied, testing whether pure and mixed types become visible.

[6] Argentine, Mexican, and Cuban novels are understood as (1) novels written by authors having that nationality or having their center of life in these countries and (2) novels first published in the countries (that might have been written by authors with another nationality). During the nineteenth century, Argentina became independent in 1816, Mexico in 1821, and Cuba only in 1898. Even though Cuba was still a colony until the end of the century, the Cuban novel developed earlier and contributed to forming a national identity (Ferrer 2018, 11–19). Nevertheless, before independence, it is most convenient to call the authors of that country Cuban-Spanish.

3. Analysis

3.1 Corpus

The corpus of novels used for the analysis includes 83 works first published between 1840 and 1910. Of these, 40 are historical (*novela histórica*), and 43 sentimental novels (*novela sentimental*). 32 of the works are Argentinian, 35 are Mexican, and 16 are Cuban. They were written by 74 different authors, 9 of them female and 65 male. Only one work per author was chosen for each subgenre to prevent the authorial signal from interfering too much with the genre signal. However, if authors wrote in both subgenres, one novel of each subgenre is included. In the corpus, there are nine authors to whom this applies. Figure 1 shows the distribution of the novels in the corpus by decade and subgenre.

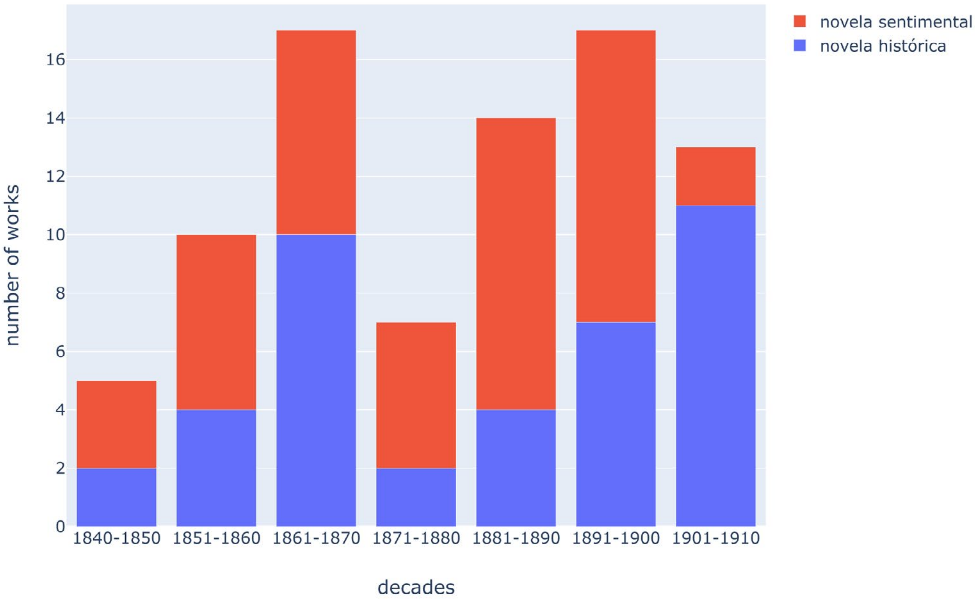

Fig. 1 Number of works in the corpus (Henny-Krahmer, CC BY).

The collection of texts used for this analysis is a subset of a larger corpus of 256 Spanish-American novels created in the project Computational Literary Genre Stylistics (CLiGS) at the University of Würzburg (Germany), which also includes novels of other subgenres. As explained below, the larger corpus is used as a basis for the generation of features.[7]

3.2 Features

It was decided to use topics as features for the analysis because they represent the themes developed in the novels and are, therefore, suitable to analyze subgenres primarily defined on a thematic basis. Topics have been tested successfully for the classification of novels by subgenre (Hettinger et al. 2016). The number of topics was set to 100, which is considered a medium degree of specification, given that the overall corpus contains 256 novels. The feature set was generated for the larger corpus with the goal of having more stable topics that represent the novel of the time in a better way than if they had been based on the smaller subcorpus. For the network analysis, only the features for the novels in the subcorpus are used. The topic model was built with the tool MALLET (McCallum 2002) and pre- and post-processed with *tmw* (Schöch and Schlör 2017). The texts were lemmatized with TreeTagger (Schmid 1995), using the Spanish parameter file, and keeping only nouns.[8]

In Figure 2, the top 40 words of four of the resulting 100 topics are visualized. They exemplify the range of themes covered in the novels. The first topic is about love and feelings: *amor* ('love'), *corazón* ('heart'), *alma* ('soul'), *pasión* ('passion'); the second topic is dominated by politics: *gobierno* ('government'), *ministro* ('minister'), *guerra* ('war'), *poder* ('power'); the third is about crime and banditry: *bandido* ('bandit'), *jefe* ('chief'), *ladrón* ('thief'), *robo* ('robbery'); and the fourth about religion and colonization: *sacerdote* ('priest'), *dios* ('god'), *español* ('Spaniard'), *guerrero* ('warrior'). The first number in parentheses indicates the rank of the topic by its probability in the whole corpus. The lower the number, the more important the topic is. Hence, the love topic

[7] The metadata of the corpus used for this analysis is available as "metadata.csv" in the folder "corpus_metadata" at https://github.com/hennyu/family_resemblance_dsrom19 and the metadata of the larger corpus as "metadata_full.csv". The whole corpus is called "Corpus de novelas hispanoamericanas del siglo XIX (conha19)" and is published at https://github.com/cligs/conha19. Besides the corpus metadata, the first GitHub repository mentioned also includes other scripts, data, and figures used in this article. Both links were accessed on May 26, 2021.

[8] In addition, a list of stop words was prepared based on the 50 most frequent nouns and adapted manually. To this, some more stop words were added manually after inspecting the results of the topic model (e.g., proper names or very general nouns). Before running the topic modeling, the texts were first lemmatized and then segmented into chunks with a length of 1000 tokens. Besides the number of topics, the topic model was created with 5,000 iterations and a hyperparameter optimization interval of 100. The feature matrices, both for the full and the reduced corpus, can be viewed on GitHub (see footnote 7).

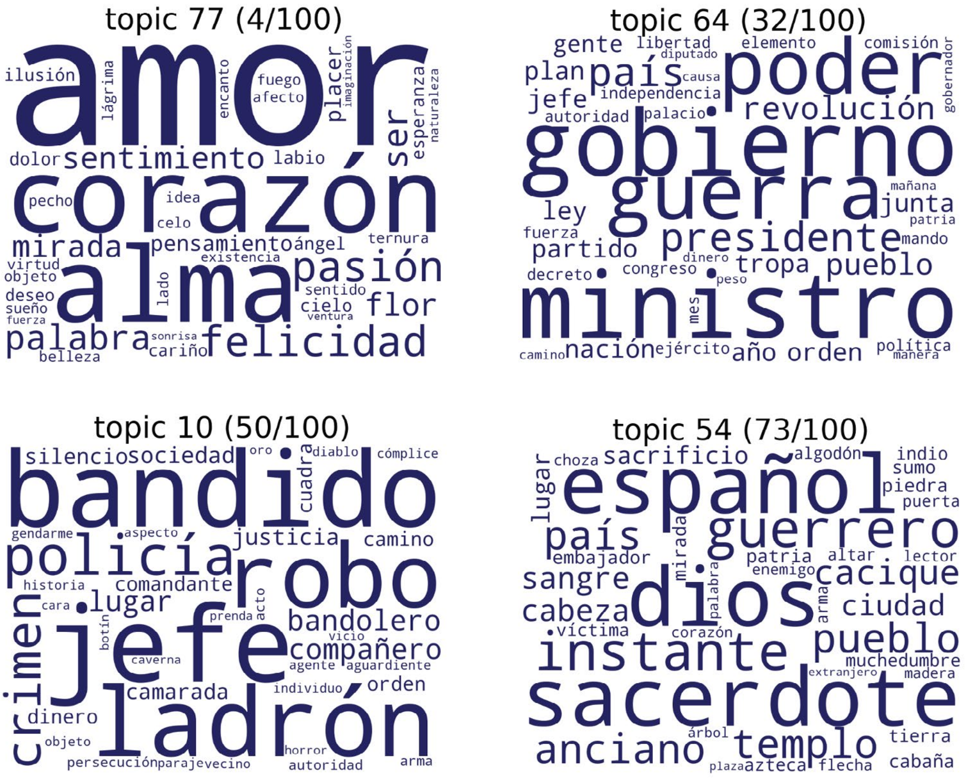

Fig. 2 Examples for topics (Henny-Krahmer, CC BY).

(rank 4) is a very general one, the politics (rank 32) and crime (rank 50) topics are still common, and the colonization topic (rank 73) is more special.

3.3 Network Analysis

To create the network for the family resemblance analysis, first, the similarities between all the individual novels were calculated for the feature set using cosine similarity.[9] After that, the resulting textual similarities were mapped onto a network structure. The nodes in the network are constituted by the novels themselves. The network relationships (or *edges*) were determined using the three nearest neighbors of each text, which

[9] Cosine similarity measures the cosine of the angle between two text vectors, see Singhal (2001).

were selected from a ranking of the text similarities. The strength (or *weight*) of the edges was calculated by summing up the similarity values of the neighbors.[10]

In the overall similarity matrix, there are relationships between all the texts. Selecting only the connections of the three nearest neighbors reduces the network's complexity and makes the closest relationships more salient. This, in turn, enhances the interpretability of the network. The choice of three is arbitrary and could be varied. However, using more than one nearest neighbor makes the results of the network more stable, as Eder (2017, 56–60) has shown. He introduced the idea of visualizing nearest neighborships based on textual similarities in a network structure to make the results of stylometric cluster analysis more reliable. This technique is adapted here with a different aim: to formalize the family resemblance concept for genre analysis.

In addition to creating the basic network structure, community detection was used to explore *families* of novels in the network. *Communities* are sets of nodes in a network that are more densely connected to each other than to nodes outside (Javed et al. 2018, 87–90). Different algorithms for the detection of network communities exist. It was decided to use the Louvain modularity algorithm (Blondel et al. 2008) because it is a suitable algorithm for detecting disjoint communities in static networks, is comparatively efficient, and has been implemented in Python, which is used to create and visualize the network here.[11]

Reflecting on how the concept of family resemblance is formulated by Wittgenstein and in literary genre theory, on the one hand, and how it is implemented here, on the other, the following observations can be made. First, using similarity relationships between the novels based on feature distributions means that it is not the presence or absence of a trait which determines the connection between members of a family and the difference to other families, but the numerical strengths of the features in combination. This transfers the idea of partial and overlapping similarities to a quantitative approach.[12] Second, when distinct communities are calculated and interpreted as families, the boundaries of the categories are sharpened retroactively. This is an advantage that balances out the looseness of the original family resemblance concept. Nevertheless, there is a significant difference between the families based on communities and conventional classes because the former emerge from a network of similarities and not from shared common features. The communities mark a boundary between one group

[10] Because the closest neighborship depends on the perspective, it was calculated for each node. When two nodes are mutually closest, the strength of the edge increases.

[11] See https://github.com/taynaud/python-louvain (accessed April 15, 2021).

[12] Of course, zero values are also possible in the feature matrices and could be interpreted as *absent*, but it would not be proportionate to consider all values that are greater than zero as *present*. A possibility to model the features in a different way would be to define a threshold value and convert all values below it to zero and all values above it to one to get a binary distinction. Still, good reasons would have to be given for the value at which to set the threshold.

of dense relationships and another, they cut off the family at a certain point, but they do not lever out the basic idea of family resemblance.[13]

3.4 Results

With the approach outlined in the previous section, three kinds of networks were produced, two for the individual subgenres and one for the two subgenres combined, as shown in Table 1.[14]

Table 1 Overview of the family resemblance networks produced.

shortcut	subgenre(s)	number of novels	number of communities (*families*)
HIST	historical novels	40	6
SENT	sentimental novels	43	6
HIST-SENT	historical and sentimental novels	83	8

The last column indicates how many *families,* that is, clusters based on the communities in the network, were produced. The number of clusters is identical for both historical and sentimental novels when they are analyzed separately. Given that the number of novels doubles when the two subgenres are combined, the number of resulting clusters does not grow proportionally, indicating that there is an overlap between the subgenres. Due to the lack of space, only some of the results can be discussed in detail in this article, and the discussion focuses on historical novels.[15] Figure 3 shows the first network for historical novels and topics (HIST). The communities detected are indicated by the different colors of the nodes.

An important question for the interpretation of the network is which kinds of novels constitute the different families. Before looking at different clusters in detail,

[13] So far, many decisions have been taken to formalize the family resemblance concept for the case study in this article. It becomes clear that variants of this approach are possible. For example, the similarity measure used, the number of nearest neighbors considered, the way to determine the strength of the edges, and the kind of community detection algorithm could be varied. As with feature-based categorization in general, also here, the selection of the features and their modeling and parametrization are subject to choice. Further empirical studies and serial analyses are needed to test the effects of such variation on the results.

[14] The script calling the various functions of the network analysis for the different setups is available at https://github.com/hennyu/family_resemblance_dsrom19/blob/main/analysis/run_scripts.py (accessed May 26, 2021).

[15] The overall results can be inspected on GitHub, though (see footnote 7), where also an analysis with the most frequent words (MFW) instead of topics is included.

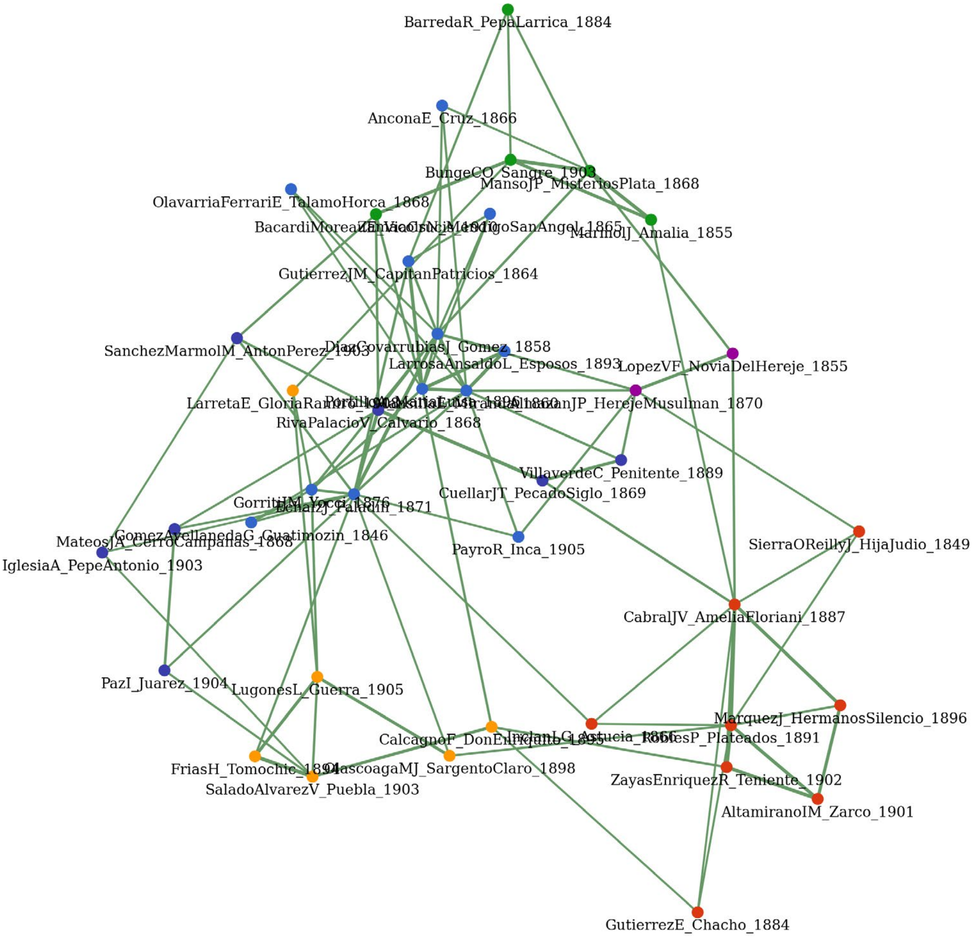
BarredaR_PepaLarrica_1884
AnconaE_Cruz_1866
BungeCO_Sangre_1903
MansoJP_MisteriosPlata_1868
OlavarriaFerrariE_TalamoHorca_1868
BacardiMoreauE_ViaCrucis_1910
ZamacoisN_MendigoSanAngel_1865
MarmolJ_Amalia_1855
GutierrezJM_CapitanPatricios_1864
SanchezMarmolM_AntonPerez_1903
DiazCovarrubiasJ_Gomez_1858
LarrosaAnsaldoL_Esposos_1893
LopezVF_NoviaDelHereje_1855
LarretaE_GloriaRamiro_1908
Portillo_MariaLuisa_1896
MansillaE_LuciaMiranda_1860
AlmazanJP_HerejeMusulman_1870
RivaPalacioV_Calvario_1868
VillaverdeC_Penitente_1889
CuellarJT_PecadoSiglo_1869
GorritiJM_Yocci_1876
EchaizJ_Paladin_1871
GomezAvellanedaG_Guatimozin_1846
MateosJA_CerroCampanas_1868
IglesiaA_PepeAntonio_1903
PayroR_Inca_1905
SierraOReillyJ_HijaJudio_1849
CabralJV_AmeliaFloriani_1887
PazI_Juarez_1904
LugonesL_Guerra_1905
MarquezJ_HermanosSilencio_1896
CalcagnoF_DonEnriquito_1895
InclanLG_Astucia_1866
RoblesP_Plateados_1891
FriasH_Tomochic_1894
OlascoagaMJ_SargentoClaro_1898
SaladoAlvarezV_Puebla_1903
ZayasEnriquezR_Teniente_1902
AltamiranoIM_Zarco_1901
GutierrezE_Chacho_1884

Fig. 3 Network of historical novels based on topics (HIST), (Henny-Krahmer, CC BY).

an overview of the cluster sizes was generated, and the possible influence of some text-external and -internal factors on the clusters was calculated, as displayed in Figures 4 and 5.

Four of the resulting six clusters are evenly sized, with 8 novels each; the other two are smaller. Cuban novels are only contained in clusters 1, 2, 3, and 5. Clusters 1, 4, and 5 are dominated by Mexican novels, and clusters 2 and 3 by Argentine novels. Cluster 3 is a Argentine–Cuban cluster, and cluster 5 a Mexican–Cuban cluster. Even

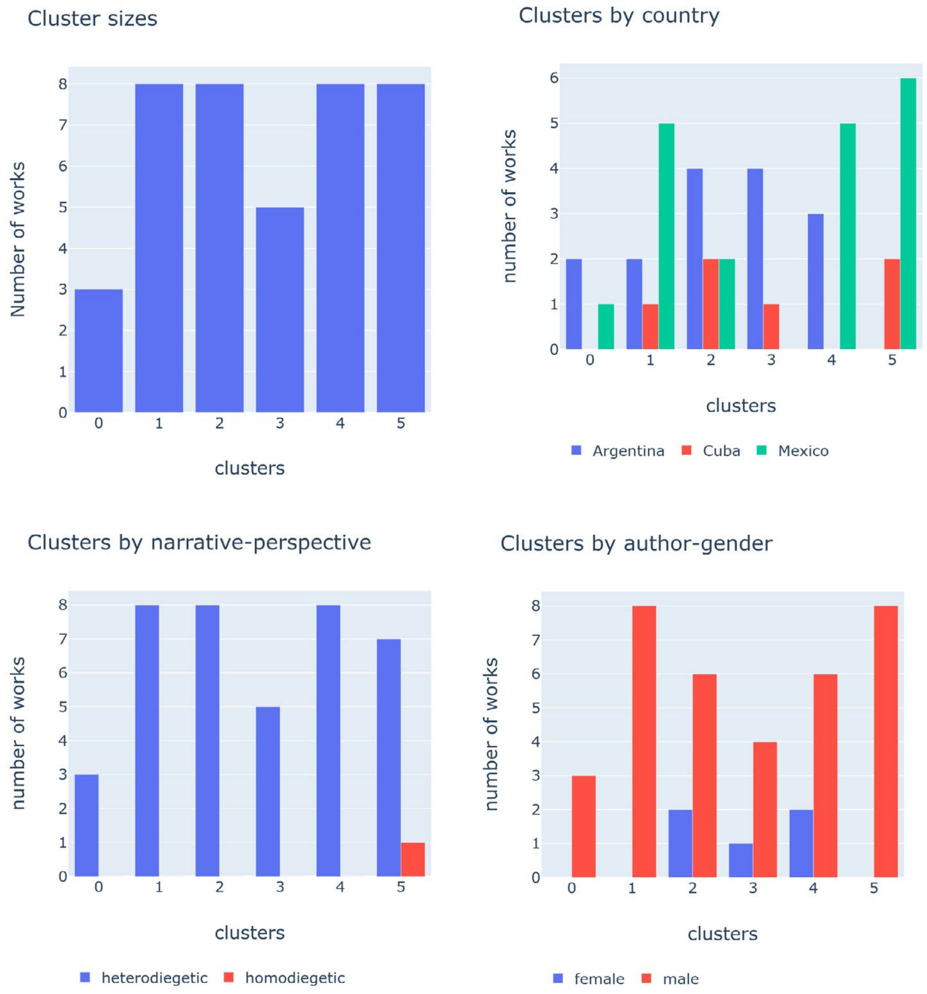

Fig. 4 Overview of cluster metadata in the network HIST (Henny-Krahmer, CC BY).

if there are some tendencies regarding the distribution of novels by country in the different clusters, there is no cluster consisting only of novels from one country. It should also be kept in mind that the overall number of novels in the individual clusters is quite small. The narrative perspective is not significant for the historical novels because there is only one novel with a homodiegetic narrator, the others all have a heterodiegetic narrator. The five historical novels written by female authors are distributed over the three clusters 2, 3, and 4, so there is no clear female cluster. Regarding the distribution

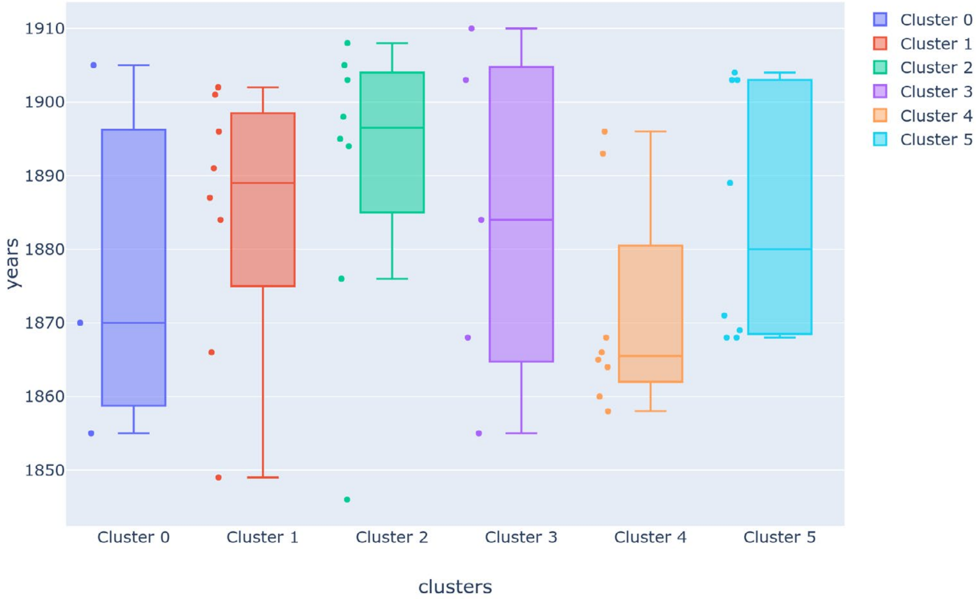

Fig. 5 Clusters by year in the network HIST (Henny-Krahmer, CC BY).

of the novels over the years, there is also much overlap, as all the clusters have earlier and later novels. Apart from one outlier, cluster 2 is rather late, and cluster 4 is mostly filled with earlier works.

Looking at one cluster in detail, it is possible to retrace the family resemblance relationships. In Table 2, the novels contained in cluster 3 are listed together with their nearest neighbors (N-1, N-2, N-3), including the weight of the edge to the respective neighbor. The strongest relationship exists between *La novela de la sangre* (Arg., 1903) by Carlos Octavio Bunge and *Los misterios del Plata* (Arg., 1868) by Juana Manso de Noronha because they are mutually closest to each other. Other bilateral nearest neighborships between novels in the cluster are highlighted in lighter orange. The novel *Pepa Larrica* (Arg., 1884) by Rafael Barreda has two nearest neighbors in the cluster, but the relationships are only unilateral. Boxes highlighted in gray show which nearest neighbors are outside of the current cluster. It becomes clear that some novels are central members of the family while others are rather distant relatives.

The topic distributions for the five historical novels are visualized in Figure 6 below to see what topics are decisive for the relationships in this cluster. The axis on the top shows the absolute value that the topic achieved in each novel, and the axis to the

Table 2 Nearest neighbors in cluster 3 of the network HIST.

idno	author	title	N-1		N-2		N-3	
nh0017	Mármol	Amalia	Misterios	1.4	Sangre	1.2	Cl 1	0.3
nh0081	Bunge	La novela de la sangre	Misterios	1.5	Crucis	1.2	Amalia	1.2
nh0094	Manso	Los misterios del Plata	Sangre	1.5	Amalia	1.4	Cl 4	0.6
nh0160	Barreda	Pepa Larrica	Cl 4	0.4	Misterios	0.4	Sangre	0.4
nh0166	Bacardí Moreau	Vía Crucis	Sangre	1.2	Cl 4	0.6	Cl 5	0.6

left shows the individual 100 topics.[16] In addition to the lines for the five novels in the cluster, a black dashed line indicating the mean topic values for all the historical novels in the network is added. The topics are ordered by importance in the whole corpus of 256 novels from top to bottom so that more general topics are at the top and more special topics are further down. Some topics of interest are labeled, the black ones being particularly important for this cluster and the red ones less important when compared to all the historical novels in the corpus.

The family approach is visible because not all decisive topics are equally relevant for the individual novels in the cluster. For example, the topics *sacerdote-dios-español* ('priest-god-Spaniard') and *fortaleza-batería-plaza* ('fortress-battery-square') are underrepresented in the whole cluster, but *amor-corazón-alma* ('love-heart-soul') and *soldado-fuego-columna* ('soldier-fire-column') are only partly less relevant. The first corresponds to the mean for the novel *Amalia*, and the second reaches almost the mean for *Vía Crucis*. Topics that are overrepresented in several novels in the cluster are *voz-palabra-brazo* ('voice-word-arm'), *idea-espíritu-instante* ('idea-spirit-moment'), *pueblo-ley-país* ('people-law-country'), *calle-puerta-voz* ('street-door-voice'), *agua-cuerpo-sangre* ('water-body-blood'), *gobierno-ministro-guerra* ('government-minister-war'), *puerta-espíritu-cabeza* ('door-spirit-head') and *cabeza-rosa-asesino* ('head-rose-assassin'). They stand for the general characteristics of the family: historical novels that are not mixed with love stories so much, not focused on military actions and not about the conquest or colonial history, but about political ideas and conditions, (inter)personal contacts and states, voices, words, and bodies. However, as specific topic values are not necessary conditions, some of the novels have their own special topics. The topic *mar-buque-puerto* ('sea-boat-harborur') is specific for *Los misterios del Plata*, *negro-esclavo-amo* ('black-slave-lord') for *Vía Crucis*, the only Cuban novel in this cluster, and *capitán-voz-revolución* ('captain-voice-revolution') for *Pepa Larrica*.

[16] In a strict sense, the topics are categories and not numerical values and should be visualized as bars rather than lines. The line plot was chosen here because it facilitates seeing the differences between the data series.

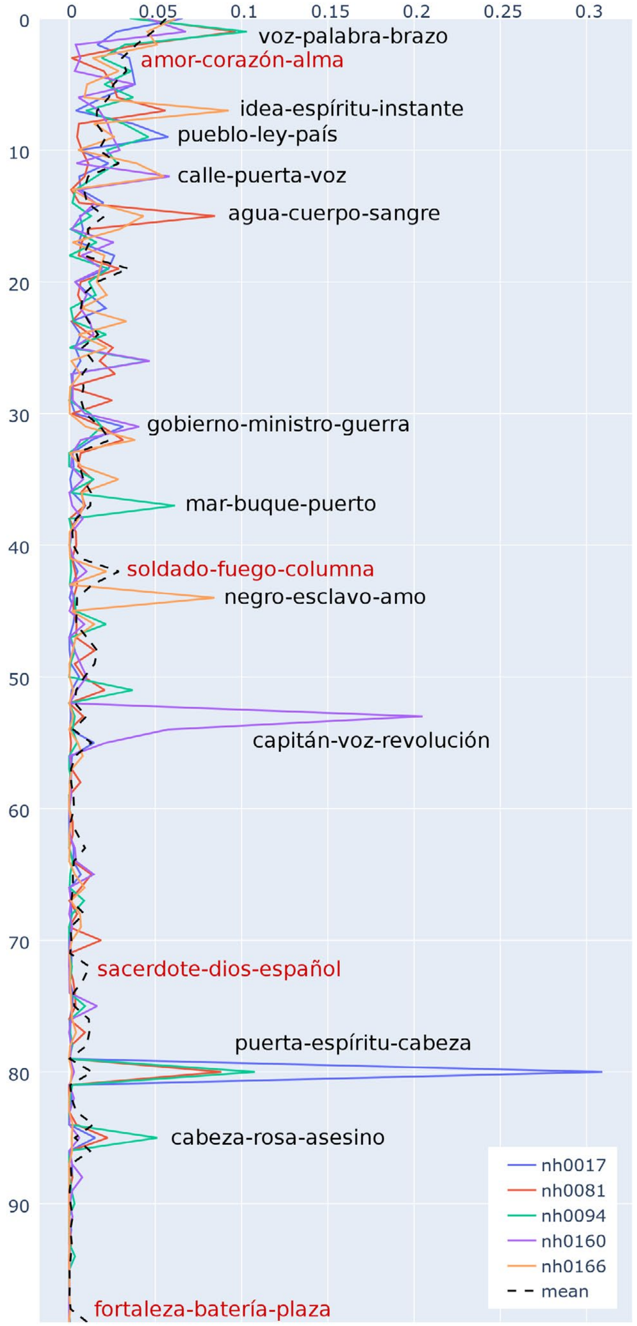

Fig. 6 Topic scores for cluster 3 in the network HIST (Henny-Krahmer, CC BY).

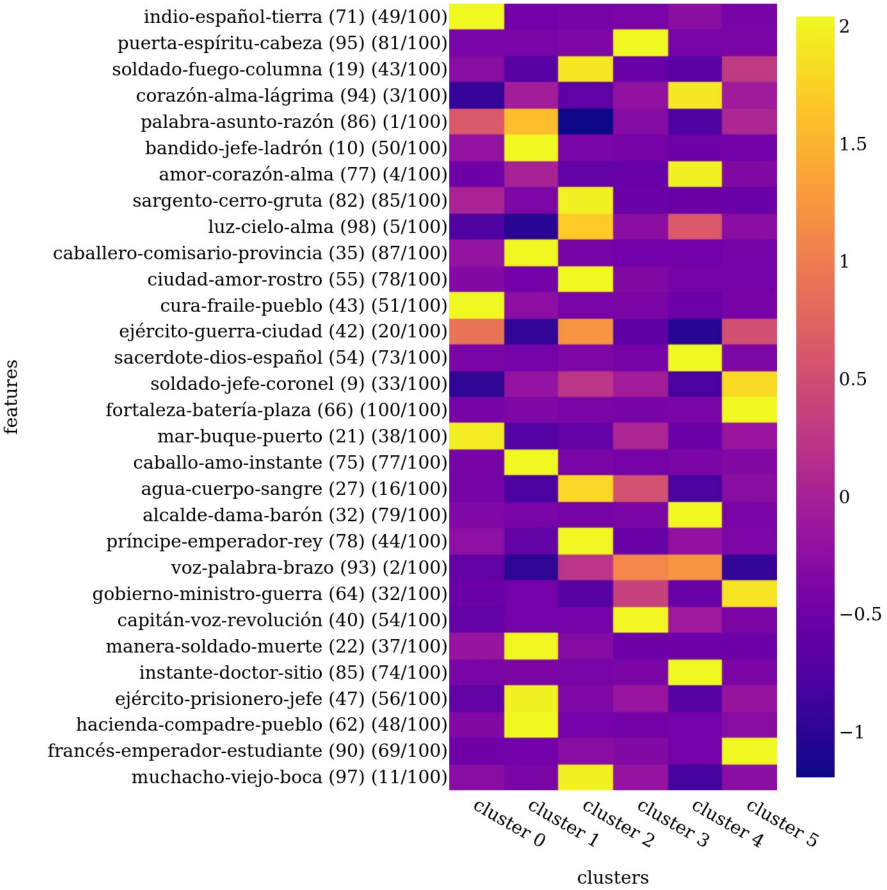

Fig. 7 Top distinctive topics in the clusters of the network HIST (Henny-Krahmer, CC BY).

A more general overview of the topics that are distinctive for the different clusters in the network of historical novels is given in Figure 7. In the heatmap, the yellower the boxes, the more important the topics are for the cluster, and the bluer, the less important they are. The distinctiveness was calculated by normalizing the topic values to z-scores (Oakes 2003, 7–8). Here, only the top 30 most distinctive topics are shown. The values in parentheses at the end of the topic labels indicate the ranks of the topics in the whole corpus (by probability), so the topic *voz-palabra-brazo* ('voice-word-arm') with rank 2, for example, is much more general than *fortaleza-batería-plaza* ('fortress-battery-square') with rank 100.

The distinctive topics of cluster 3 that were already discussed can be recognized in the heatmap. The smallest cluster 0 seems to be about the conquest and colonial history, as the most distinctive topics are *indio-español-tierra* ('Indian-Spaniard-land'), *cura-fraile-pueblo* ('priest-friar-village'), and *mar-buque-puerto* ('sea-ship-harbor'). In cluster 1, topics about military campaigns and rural life prevail, making one think about internal struggle, bandits and *gauchos*: *palabra-asunto-razón* ('word-matter-reason'), *bandido-jefe-ladrón* ('bandit-chief-thief'), *caballero-comisario-provincia* ('gentleman-inspector-province'), *caballo-amo-instante* ('horse-lord-moment'), *manera-soldado-muerte* ('manner-soldier-dead'), *ejército-prisionero-jefe* ('army-prisoner-chief'), *hacienda-compadre-pueblo* ('estate-godfather-village'). Cluster 2 is not so easy to interpret. It is about military action (*soldado-fuego-columna* ('soldier-fire-column'), *sargento-cerro-gruta* ('sargeant-hill-grot'), *ejército-guerra-ciudad* ('army-war-city'), but there are other, individual topics. Cluster 4 is clearly romantic with colonial and aristocratic elements: *corazón-alma-lágrima* ('heart-soul-tear'), *amor-corazón-alma* ('love-heart-soul'), *sacerdote-dios-español* ('priest-god-Spaniard'), *alcalde-dama-barón* ('mayor-lady-baron'), *instante-doctor-sitio* ('moment-doctor-place'). This fits well with the observation that it contains mostly earlier novels published during the main phase of the Romantic current in Spanish-America in the first half of the nineteenth century. The last cluster is politico-historical with the top topics *soldado-jefe-coronel* ('soldier-chief-colonel'), *fortaleza-batería-plaza* ('fortress-battery-square'), *gobierno-ministro-guerra* ('government-minister-war'), and *francés-emperador-estudiante* ('French-emperor-student').

For reasons of space, the results for the network of sentimental novels and the one containing both types of subgenres are only summarized briefly here. Regarding the metadata, the cluster sizes vary more for the sentimental novels, with the biggest cluster having 14 novels and the smallest one only two. All the clusters are mixed by country. Among the sentimental novels, there are more with autodiegetic and homodiegetic narrators, and the narrative perspective has an influence on the results. The smallest cluster, for instance, consists solely of autodiegetic texts featuring topics related to inner life and landscape. For the sentimental novels, there are also clearer tendencies of topic changes over time, as Figure 8 shows. The early cluster 0 is romantic with letters, dance, aristocracy, and much emotionality. The three later clusters are the ones dominated by interiorization, and the mid-century cluster 5 is worldly about food, marriage, business, and money.

The most important point to note when both subgenres are analyzed together is that the subgenres are not neatly sorted into the different families. As can be seen in Figure 9, there are clusters dominated by one subgenre—clusters 1, 3, and 6 by sentimental novels and clusters 2, 4, and 7 by historical novels—but there is no cluster containing only novels of one subgenre. The clusters 0 and 5 are entirely mixed. This indicates that it could be difficult to classify these novels by the primary type to which they have been assigned by their authors and by critics.

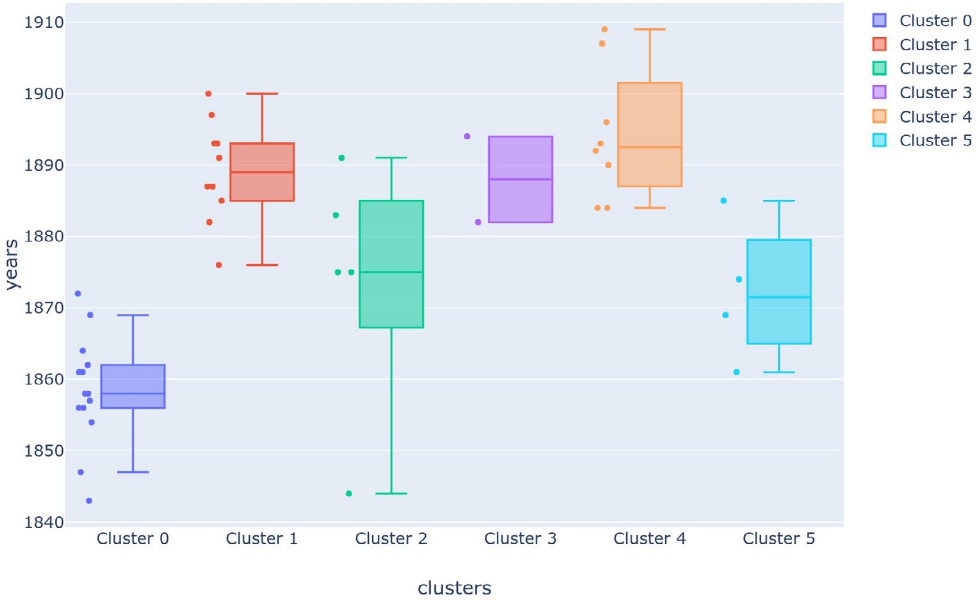

Fig. 8 Clusters by year in the network SENT (Henny-Krahmer, CC BY).

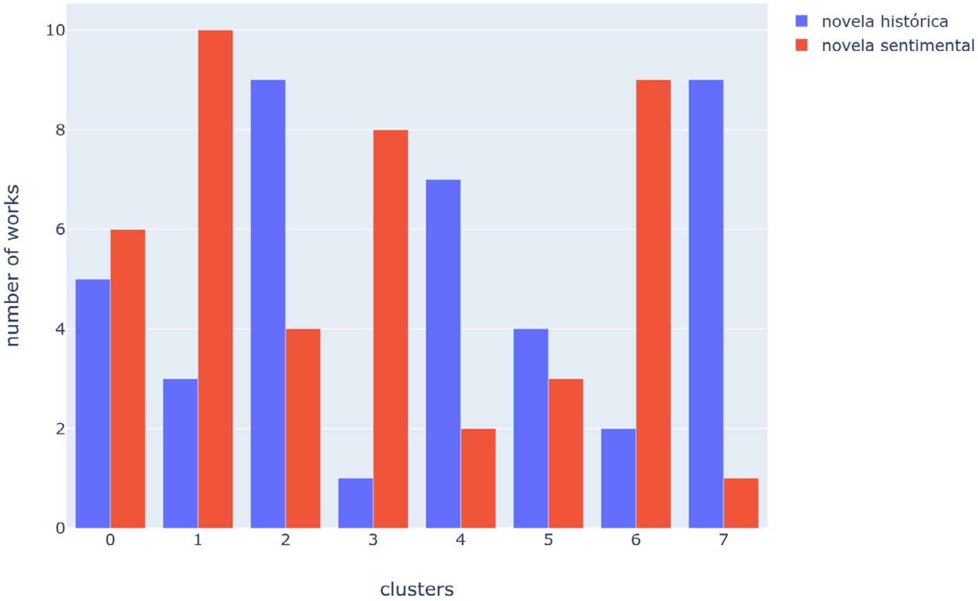

Fig. 9 Clusters by subgenre in the combined network (Henny-Krahmer, CC BY).

In the combined network, the cluster sizes vary moderately from seven to 13 novels. Here again, there is no clear tendency for countries. Different narrative perspectives are not concentrated in single clusters, so this aspect, which was observed for the sentimental novels alone, disappears when they are analyzed in the more general setup. Regarding the distribution by years, cluster 1 is early, clusters 3, 4, and 6 are late, and the others are mixed. The topics that are distinctive for the different families reflect the relative purity or mixture of subgenres as well as the preferences of the early versus the late nineteenth century.

4. Conclusion

Here a proposal was made for how the concept of family resemblance, which was introduced into genre theory in the 1960s and also argued for by several genre theorists recently, can be applied in a digital genre stylistics approach. With the analysis of topics in nineteenth-century Spanish-American historical and sentimental novels, the proposal was empirically tested in a network-based approach. If one looks at the current strategies to categorize genres in digital stylistics, the majority focus on classificatory groupings based on the assumption of features that are common to all members of a class. However, there are also alternative ways to analyze genres in digital stylistics, some of which have explicitly addressed genre theoretical questions, while others have not. In particular, stylometric network analyses implicitly contain the idea of overlapping similarities and unsharp boundaries, which is characteristic of the family resemblance approach. In this article these two scenarios were brought together. With the chosen approach of comparing the feature distributions of the novels and organizing the resulting network of similarities into communities interpreted as families, the original idea of family resemblance is adapted for digital analysis. First, rather than the presence or absence of individual textual features, the degree of their joint presence is decisive. Second, communities or clusters found in the similarity network constitute a way to delimit the families retroactively without changing the underlying concept of intertwining shared characteristics of individual members of the groups.

For the Argentine, Mexican, and Cuban historical and sentimental novels, the analysis confirmed that there are subtypes of the subgenres that have already been described in literary historical approaches. Such subtypes are, for example, a novel with a historical setting and a sentimental plot, or a historical novel focusing on contemporary political conditions, or a historical novel about events of colonial times. In addition, influences of the narrative perspective on subtypes of the sentimental novel became visible. Analyzing both types of subgenres together resulted in mixed groups and some that are dominated by one subgenre. While the country in which the novels were published or the authors' nationalities do not have a clear impact on the resulting families of novels, the year of publication does, in some cases, when the preferred and avoided topics reflect the literary development in the nineteenth century. All in all, the results show that features common to all novels of a subgenre cannot be expected and that the factors that influence the subgroups or families of subgenres are diverse. There is not one decisive factor, each family has its own traits that hold it together, and inside of each one, there are additional individual traits as well as connections to other families.

To conclude, the algorithm producing the family resemblance network and the resulting data offer an empirical ground on which literary historians can look for sense in genre historical terms. It does not say anything about the historico-cultural and communicative relevance of the connections, but it might reveal previously unrecognized textual similarities in addition to confirming known ones on a broader textual basis. By not presupposing strict uniformity inside and strict boundaries between the genre categories, it comes closer to the open genres that the novel and its subgenres form in terms of theme and style.

ORCID®

Ulrike Henny-Krahmer https://orcid.org/0000-0003-2852-065X

References

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Calvo Tello, José. 2018. "Genre Classification in Novels: A Hard Task for Humans and Machines?" Paper presented at *EADH 2018: Data in Digital Humanities, Galway, 7–9 December 2018*. Galway: National University of Ireland. https://eadh2018.exordo.com/programme/presentation/82 (Accessed April 15, 2021).

Calvo Tello, José, Daniel Schlör, Ulrike Henny, and Christof Schöch. 2017. "Neutralizing the Authorial Signal in Delta by Penalization: Stylometric Clustering of Genre in Spanish Novels." In *Digital Humanities 2017. Conference Abstracts, Montréal, Canada, 8–11 August 2017*, 181–184. Montréal: McGill University and Université de Montréal.

Dill, Hans-Otto. 1999. *Geschichte der lateinamerikanischen Literatur im Überblick*. Stuttgart: Reclam.

Eder, Maciej. 2017. "Visualization in Stylometry. Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32 (1). https://doi.org/10.1093/llc/fqv061.

Ferrer, José Luis. 2018. *La invención de Cuba: Novela y nación (1837–1846)*. Madrid: Editorial Verbum.

Fishelov, David. 1993. *Metaphors of Genre: The Role of Analogies in Genre Theory*. University Park, PA: Pennsylvania State University Press.

Fludernik, Monika. 2009. “Roman.” In *Handbuch der literarischen Gattungen*, edited by Dieter Lamping and Sandra Poppe, 627–45. Stuttgart: Kröner.

Fowler, Alastair. 1982. *Kinds of Literature. An Introduction to the Theory of Genres and Modes*. Oxford: Clarendon Press.

Fricke, Harald. 2010. “Definitionen und Begriffsformen.” In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 7–10. Stuttgart, Weimar: J.B. Metzler.

Gálvez, Marina. 1990. *La novela hispanoamericana (hasta 1940)*. Madrid: Taurus.

Gianitsos, Efhimios Tim, Thomas J. Bolt, Pramit Chaudhuri, and Joseph P. Dexter. 2019. “Stylometric Classification of Ancient Greek Literary Texts by Genre.” In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Minneapolis, MN, USA, 7 June, 2019*, 52–60. Minneapolis: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2507.

Hempfer, Klaus W. 2014. “Some Aspects of a Theory of Genre.” In *Linguistics and Literary Studies: Interfaces, Encounters, Transfers*, edited by Monika Fludernik and Daniel Jacob, 405–22. Berlin: De Gruyter.

Henny-Krahmer, Ulrike. 2018. “Exploration of Sentiments and Genre.” In *Digital Humanities 2018. Puentes–Bridges. Book of Abstracts, Mexico City, 26–29 June 2018*, 399–403. Mexico City: Red de Humanidades Digitales.

Henny-Krahmer, Ulrike, Katrin Betz, Daniel Schlör, and Andreas Hotho. 2018. “Alternative Gattungstheorien. Das Prototypenmodell am Beispiel hispanoamerikanischer Romane.” In *DHd 2018. Kritik der digitalen Vernunft. Konferenzabstracts, Köln, 26 February–2 March 2018*, 105–12. Köln: Universität zu Köln.

Hettinger, Lena, Isabella Reger, Fotis Jannidis, and Andreas Hotho. 2016. “Classification of Literary Subgenres.” In *DHd2016. Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts, Leipzig, 7–12 March 2016*, 160–64. Duisburg: nisaba verlag.

Javed, Muhammad Aqib, Muhammad Shahzad Younis, Siddique Latif, and Adeel Baig. 2018. “Community Detection in Networks: A Multidisciplinary Review.” *Journal of Network and Computer Applications* 108: 87–111. https://doi.org/10.1016/j.jnca.2018.02.011.

Lindstrom, Naomi. 2004. *Early Spanish American Narrative*. Austin: University of Texas Press.

McCallum, Andrew Kachites. 2002. “MALLET: A Machine Learning for Language Toolkit.” http://mallet.cs.umass.edu (Accessed April 15, 2021).

Molina, Hebe Beatriz. 2011. *Como crecen los hongos. La novela argentina entre 1838 y 1872*. Buenos Aires: Teseo.

Müller, Ralph. 2010. “Kategorisieren.” In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 21–23. Stuttgart, Weimar: J. B. Metzler.

Neumann, Birgit, and Ansgar Nünning. 2007. “Einleitung: Probleme, Aufgaben und Perspektiven der Gattungstheorie und Gattungsgeschichte.” In *Gattungstheorie und Gattungsgeschichte*, edited by Marion Gymnich, Birgit Neumann, and Ansgar Nünning, 1–28. Trier: WVT.

Oakes, Michael P. 2003. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Read, John Lloyd. 1939. *The Mexican Historical Novel*. New York: Instituto de las Españas en los Estados Unidos.

Schmid, Helmut. 1995. "Probabilistic Part-of-Speech Tagging Using Decision Trees." *Proceedings of the International Conference on New Methods in Language Processing, Manchester, UK, 1994*. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf. (Accessed April 15, 2021).

Schöch, Christof. 2017. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly* 11 (2). http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html (Accessed April 15, 2021).

Schöch, Christof. 2013. "Fine-tuning Our Stylometric Tools: Investigating Authorship and Genre in French Classical Theater." In *Digital Humanities 2013. Conference Abstracts, Lincoln, USA, 16–19 July 2013*, 383–86. Lincoln: Center for Digital Research in the Humanities.

Schöch, Christof, and Daniel Schlör. 2017. "tmw – Topic Modeling Workflow." *GitHub*. https://github.com/cligs/tmw (Accessed April 15, 2021).

Schöch, Christof, Ulrike Henny, José Calvo Tello, Daniel Schlör, and Stefanie Popp. 2016. "Topic, Genre, Text. Topics im Textverlauf von Untergattungen des spanischen und hispanoamerikanischen Romans (1880–1930)." In *DHd2016. Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts, Leipzig, 7–12 March 2016*, 235–39. Duisburg: nisaba verlag.

Schröter, Julian. 2024. "Machine-Learning as a Measure of the Conceptual Looseness of Disordered Genres: Studies on German *Novellen*." In *Digital Stylistics in Romance Studies and Beyond*, edited by Robert Hesselbach, José Calvo Tello, Ulrike Henny-Krahmer, Christof Schöch, and Daniel Schlör, 173–195. Heidelberg: heiUP.

Singhal, Amit. 2001. "Modern Information Retrieval: A Brief Overview." In *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (4): 35–43. http://sites.computer.org/debull/A01dec/singhal.ps (Accessed April 15, 2021).

Strube, Werner. 1993. *Analytische Philosophie der Literaturwissenschaft. Untersuchungen zur literaturwissenschaftlichen Definition, Klassifikation, Interpretation und Textbewertung*. Paderborn: Schöningh.

Tophinke, Doris. 1997. "Zum Problem der Gattungsgrenze – Möglichkeiten einer prototypentheoretischen Lösung." In *Gattungen mittelalterlicher Schriftlichkeit*, edited by Barbara Frank, Thomas Haye, and Doris Tophinke, 161–82. Tübingen: Narr.

Underwood, Ted. 2015. *Understanding Genre in a Collection of a Million Volumes*. White Paper Report. Urbana,-Champaign: University of Illinois. http://dx.doi.org/10.17613/M6W07V.

Underwood, Ted. 2016. "The Life Cycles of Genres." *Journal of Cultural Analytics*. 2 (2) https://doi.org/10.22148/16.005.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: The University of Chicago Press.

Wittgenstein, Ludwig. 2009. *Philosophical Investigations*. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Edited by P. M. S. Hacker and Joachim Schulte. New York: Wiley.