CLS INFRA
COMPUTATIONAL LITERARY STUDIES INFRASTRUCTURE

Network analysis in the humanities

Daniil Skorinkin, University of Potsdam / CLS INFRA

Our plan for today

- Network analysis: why should I use it and what for? (30 minutes)
  - Round of questions
- Technicalities of network data preparation: how do I convert my object of research to network? (20 minutes)
  - Small hands-on with EzLinaVis (a web-based tool)

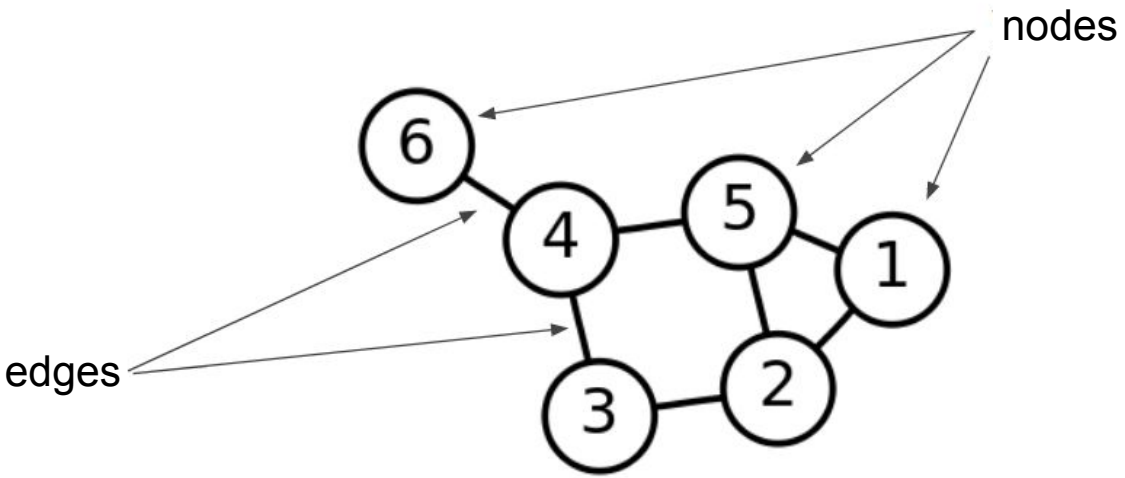
BREAK

- Technicalities of network analysis: how do I analyze my networks (rest of time)
  - Main hands-on with Gephi and real network data

Network analysis basics

Graph theory and network science

A graph (a network) in a nutshell:



First suggested by Leonard Euler in early 18 century to solve a certain transportation problem:

2. Problema autem hoc, quod mihi satis notum esse perhibebatur, erat sequens: Regiomonti in Borussia esse insulam *A*, *der Kneiphof* dictam, fluviumque eam cingentem in duos dividi ramos, quemadmodum ex figura (Fig. 1) videre licet; ramos vero huius fluvii septem instructos esse pontibus *a*, *b*, *c*, *d*, *e*, *f* et *g*. Circa hos pontes iam ista proponebatur quaestio, num quis cursum ita instituere queat, ut per singulos pontes semel et non plus quam semel transeat. Hocque fieri posse, mihi dictum est, alios negare alios dubitare; neminem vero affirmare. Ego ex hoc mihi sequens maxime generale formavi problema: quaecunque sit fluvii figura et distributio in ramos atque quicunque fuerit numerus pontium, invenire, utrum per singulos pontes semel tantum transiri queat an vero secus.

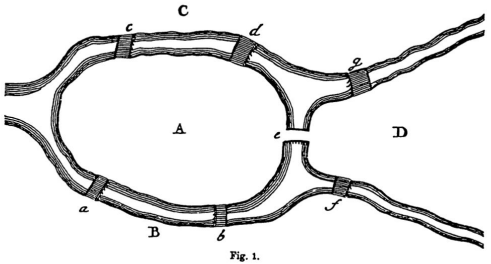
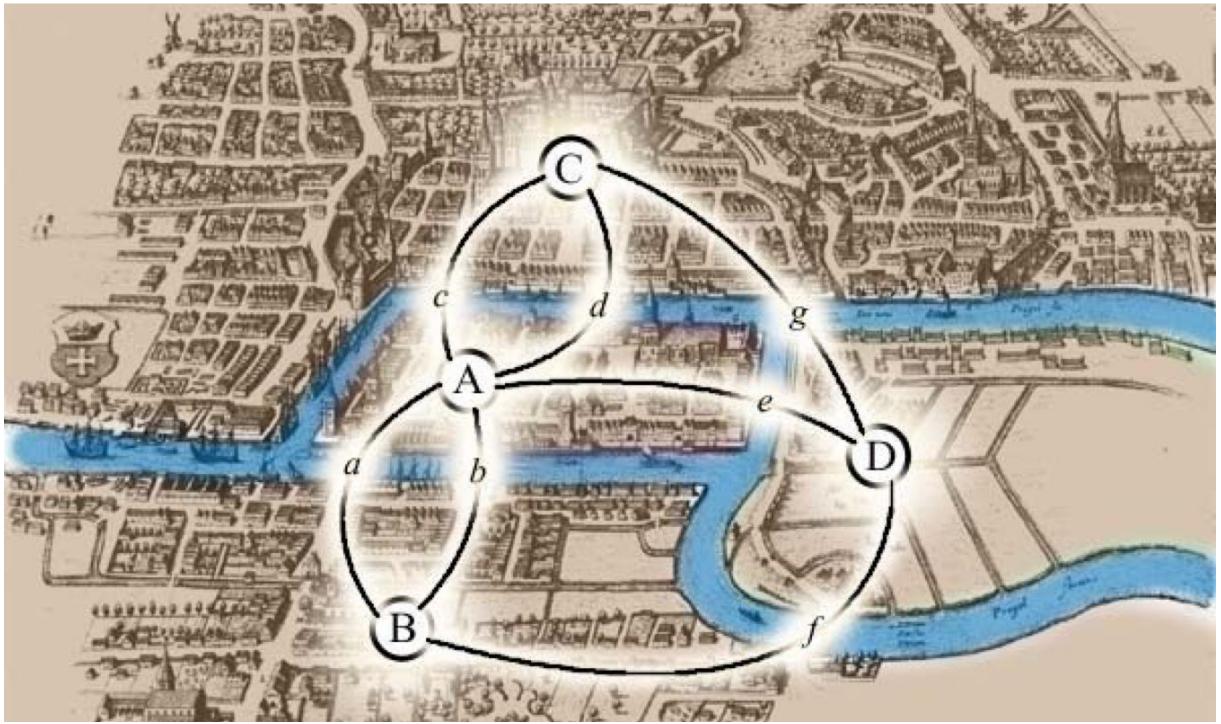


Fig. 1.

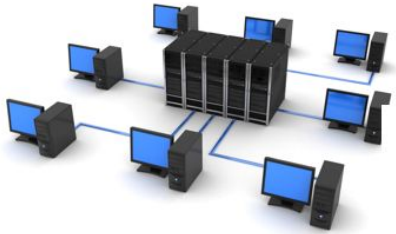
3. Quod quidem ad problema Regiomontanum de septem pontibus attinet, id resolvi posset facienda perfecta enumeratione omnium cursuum, qui institui possunt; ex his enim innotesceret, num quis cursus satisfaceret an vero nullus. Hic vero solvendi modus propter tantum combinationum numerum et nimis esset difficilis atque operosus et in aliis quaestionibus de multo pluribus pontibus ne quidem adhiberi posset. Hoc porro modo si operatio ad finem perducatur, multa inveniuntur, quae non erant in quaestione; in quo procul dubio tantae difficultatis causa consistit. Quamobrem missa hac me-

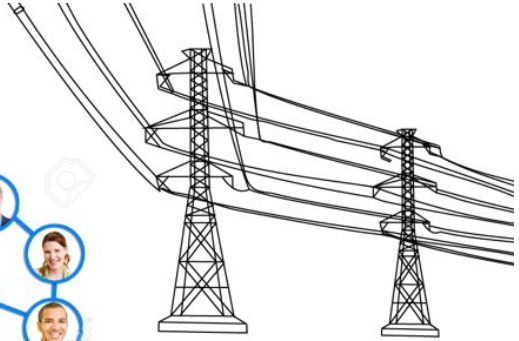
Leonhardus Eulerus. Solutio problematis ad geometriam situs pertinentis / Commentarii academiae scientiarum Petropolitanae. 8 (1736). 1741. P. 128

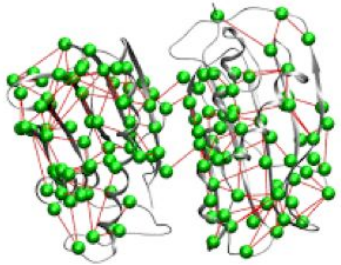
The seven bridges of Königsberg



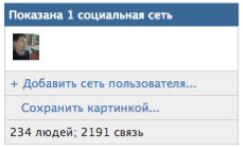
A graph could be used to model a wide range of systems

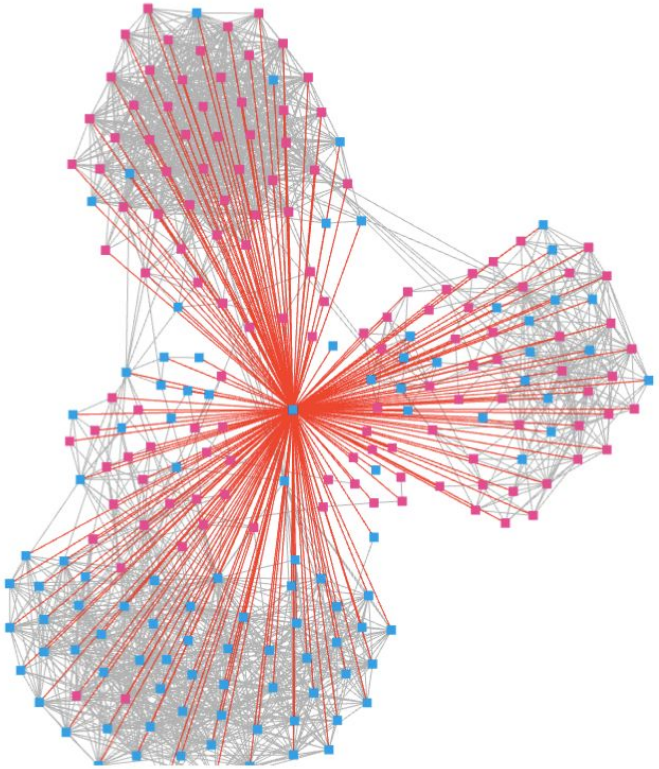






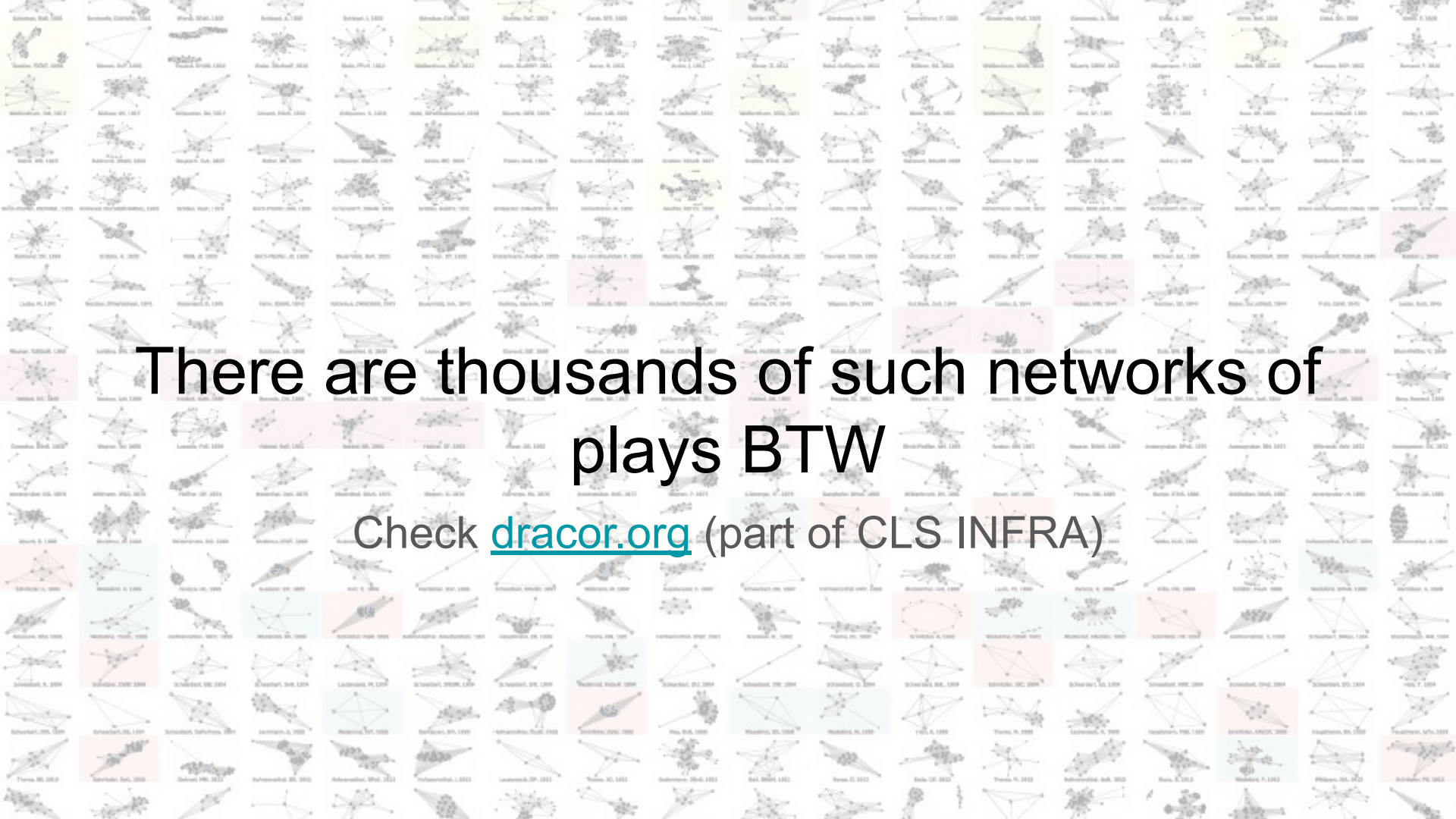
Including social systems, of course





…and even fictional social networks

https://dracor.org/ger/goethe-faust-eine-tragoedie

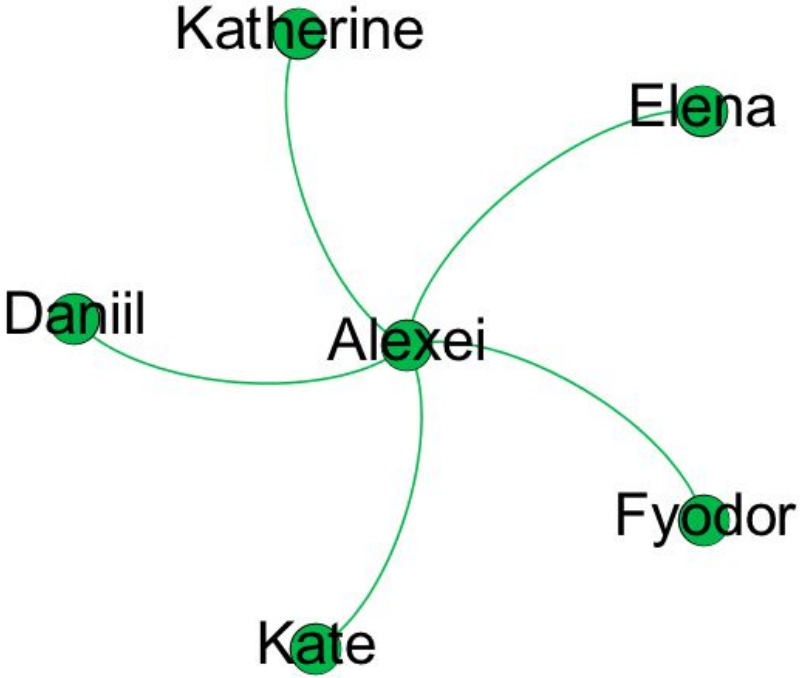


There are thousands of such networks of plays BTW

Check dracor.org (part of CLS INFRA)

Why is this model useful for us?

Reason 1: Network centralities



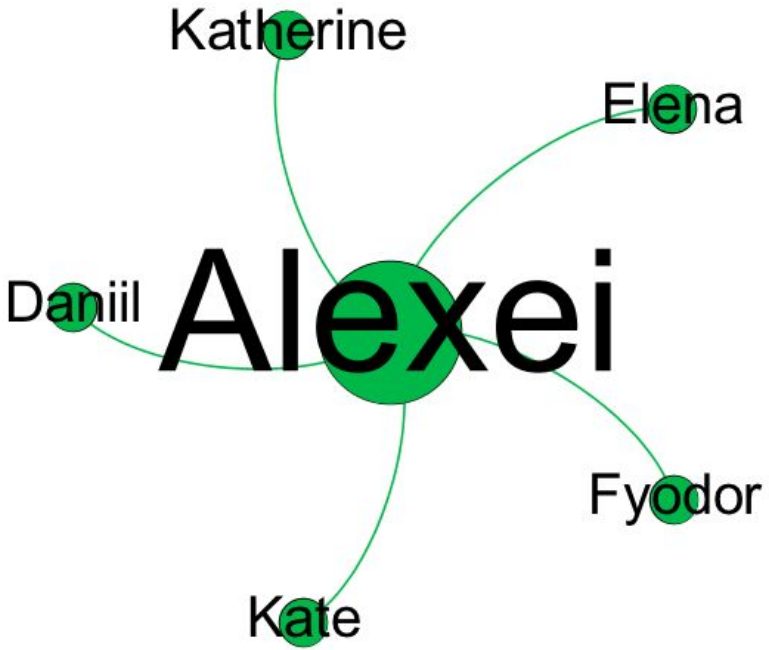
Let’s put it into numbers:

Node	Nr of connections
Alexei	5
Daniil	1
Katherine	1
Fyodor	1
Elena	1
Kate	1

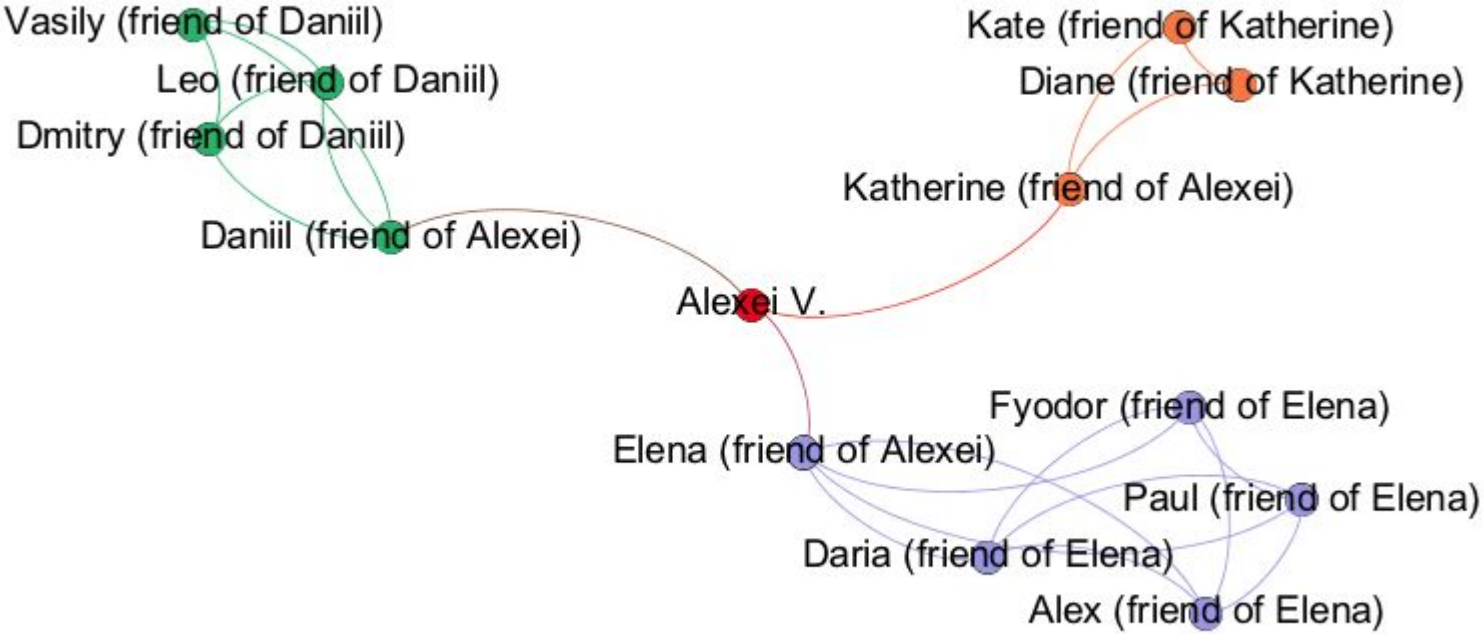
Let’s put it into numbers:

Node	**Degree centrality**
Alexei	5
Daniil	1
Katherine	1
Fyodor	1
Elena	1
Kate	1

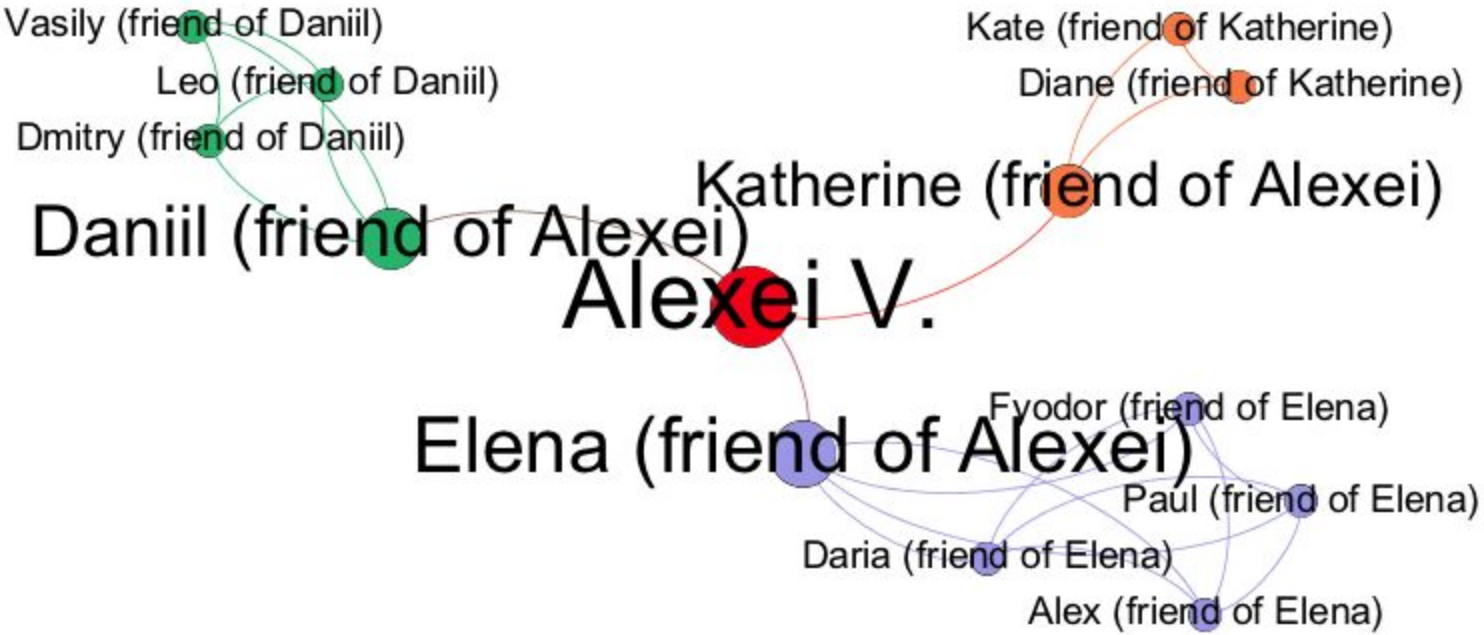
Visualizing degree centrality of this network through node size:



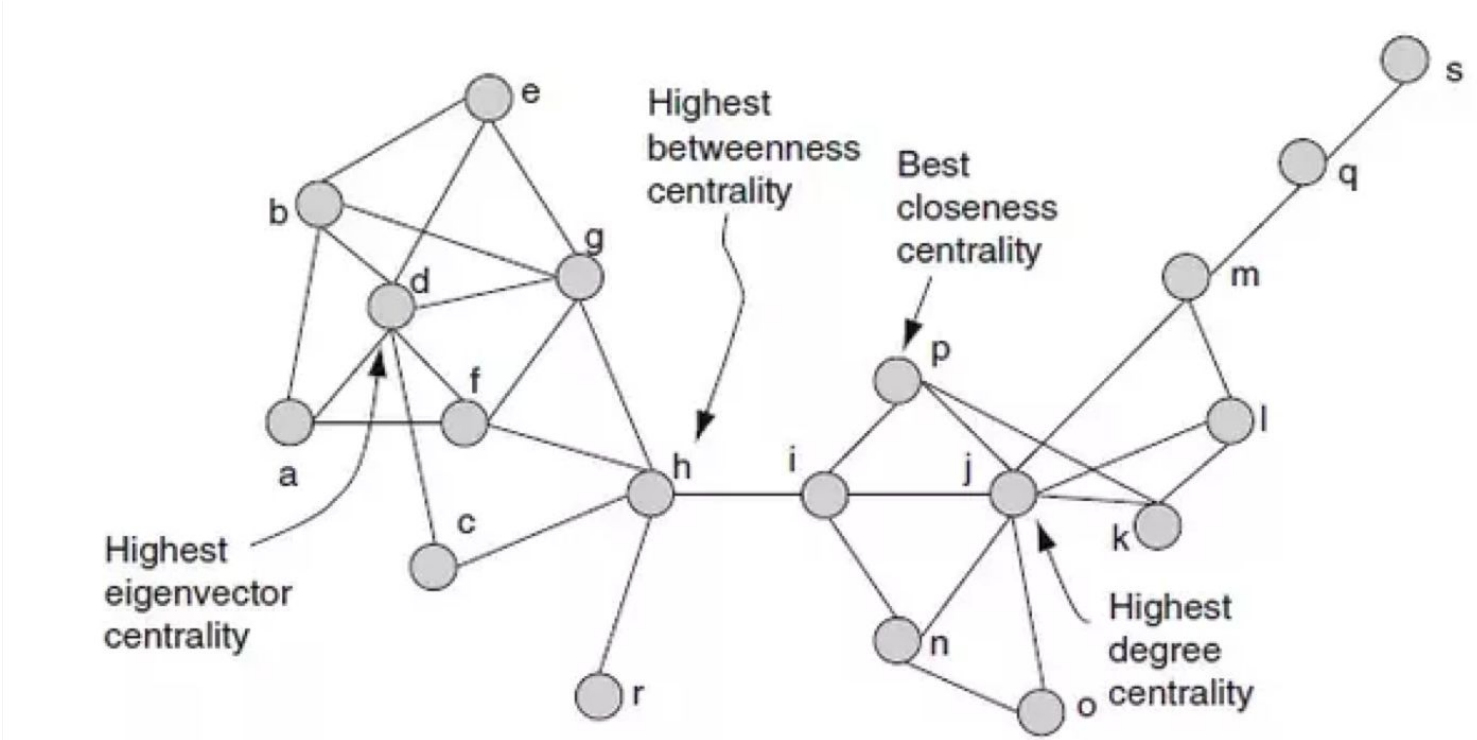
Betweenness centrality:



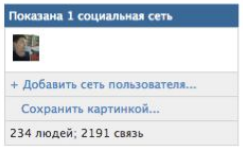
Node size proportional to betweenness:

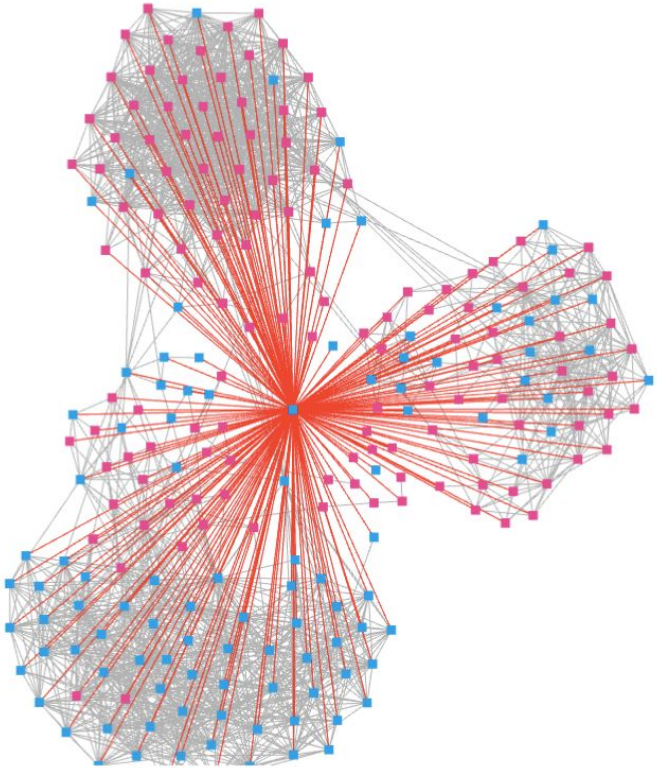


There are many more centralities which reflect different specifics of node's position in the network:

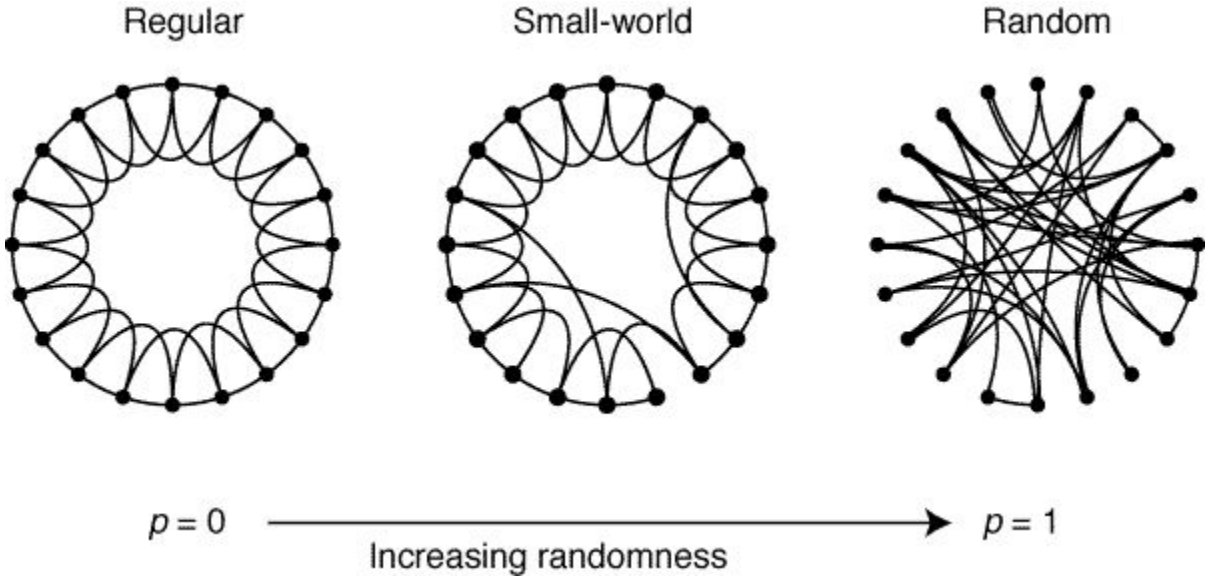


Reason 2: Subnetworks, communities, and overall structure:



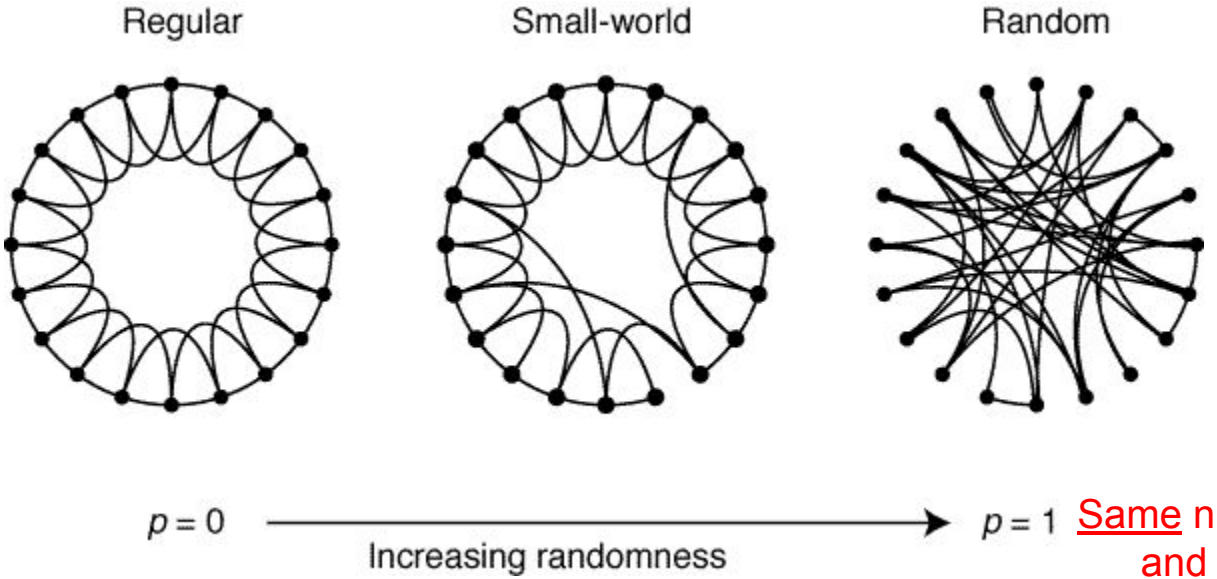


Network structures can vary:



Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918

Network structures can vary:



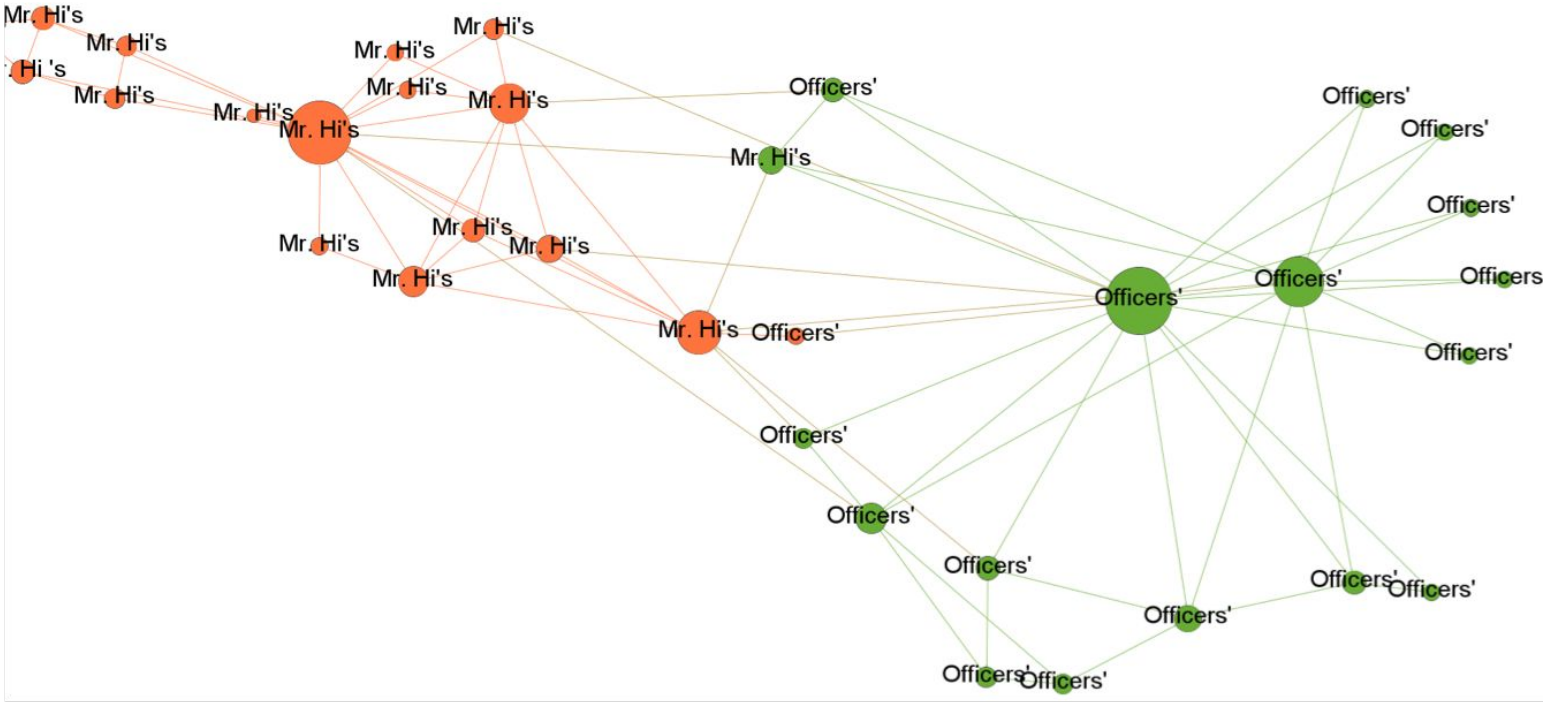
Same number of nodes and edges here!

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918

A textbook example of communities detection from social network analysis: the Karate Club (1972)

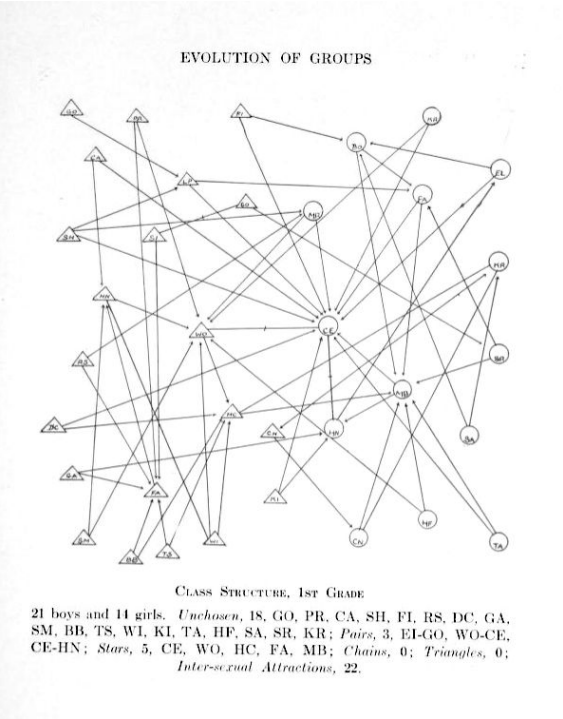
- Anthropologist Wayne Zachary studied a small karate club from 1970 to 1972
- He recorded informal connections between members outside the club
- During the study, a conflict arose between the leadership and the instructor
- Half of the members left with the instructor to start a new club
- The other half stayed (or quit karate altogether)
- Zachary wrote a program that, based on data collected before the split, tries to predict the boundaries of the group's split

My visualisation of the Karate Club network with automatic clustering of the network:



How it network analysis came to the humanities

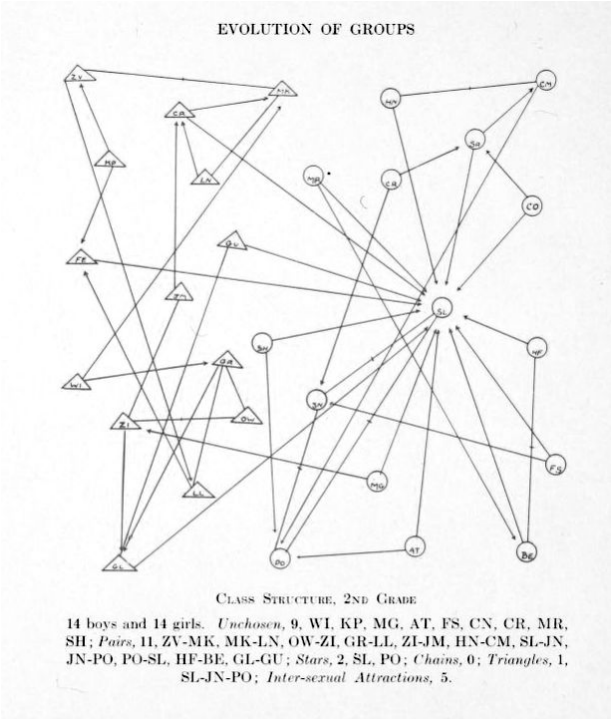
Network analysis has a long tradition in social sciences



EVOLUTION OF GROUPS

CLASS STRUCTURE, 1ST GRADE

21 boys and 14 girls. *Unchosen*, 18, GO, PR, CA, SH, FI, RS, DC, GA, SM, BB, TS, WI, KI, TA, HF, SA, SR, KR; *Pairs*, 3, EI-GO, WO-CE, CE-HN; *Stars*, 5, CE, WO, HC, FA, MB; *Chains*, 0; *Triangles*, 0; *Inter-sexual Attractions*, 22.



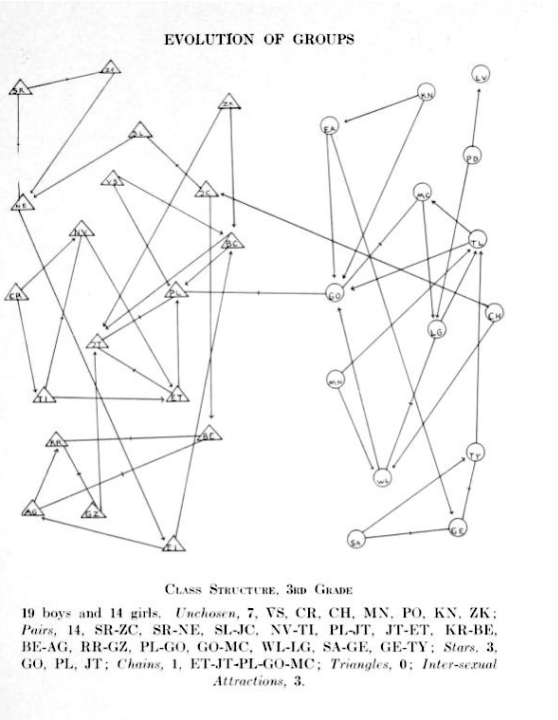
EVOLUTION OF GROUPS

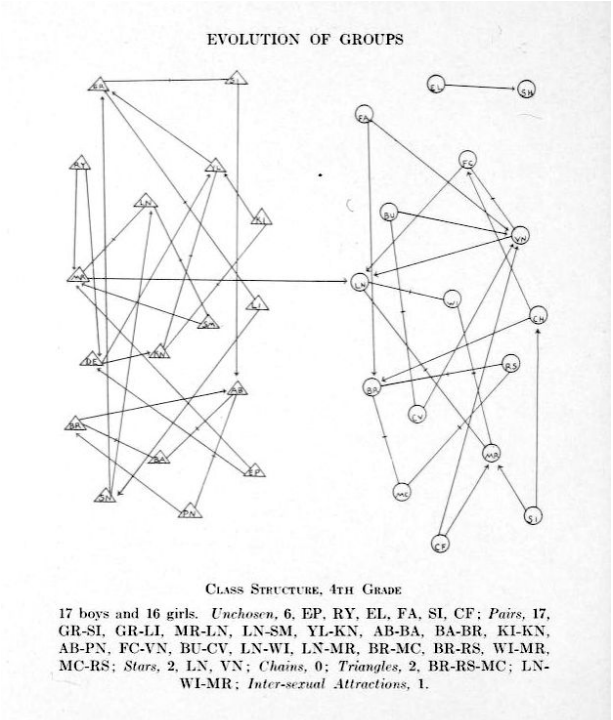
CLASS STRUCTURE, 2ND GRADE

14 boys and 14 girls. *Unchosen*, 9, WI, KP, MG, AT, FS, CN, CR, MR, SH; *Pairs*, 11, ZV-MK, MK-LN, OW-ZI, GR-LL, ZI-JM, HN-CM, SL-JN, JN-PO, PO-SL, HF-BE, GL-GU; *Stars*, 2, SL, PO; *Chains*, 0; *Triangles*, 1, SL-JN-PO; *Inter-sexual Attractions*, 5.

Moreno J. Who Shall Survive: A New Approach to the Problem of Human Interrelations. Washington: Nervous and Mental Disease Publishing Co. 1934

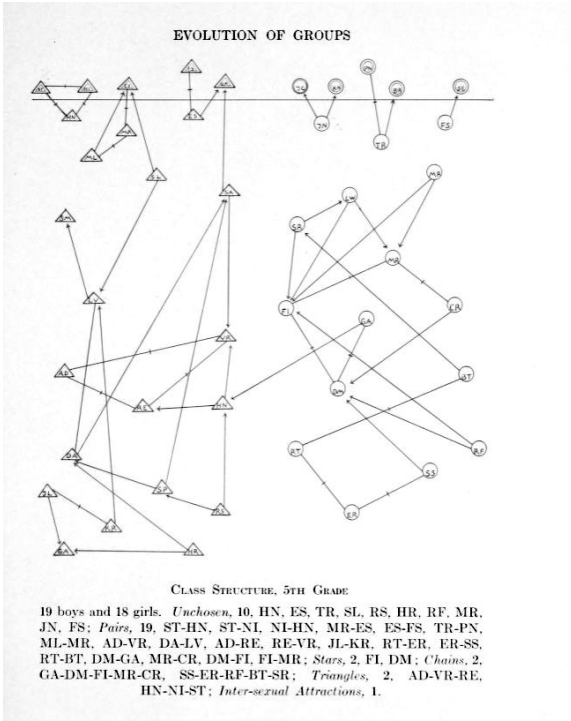
Network analysis has a long tradition in social sciences





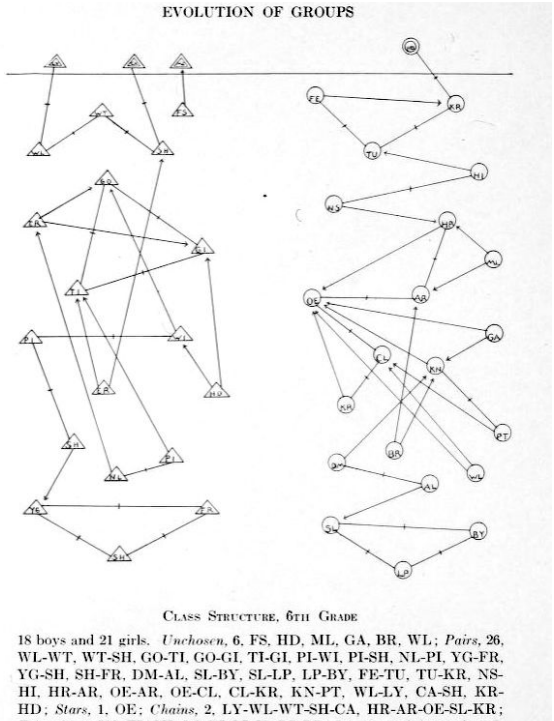
Moreno J. Who Shall Survive: A New Approach to the Problem of Human Interrelations. Washington: Nervous and Mental Disease Publishing Co. 1934

Network analysis has a long tradition in social sciences



CLASS STRUCTURE, 5TH GRADE

19 boys and 18 girls. *Unchosen*, 10, HN, ES, TR, SL, RS, HR, RF, MR, JN, FS; *Pairs*, 19, ST-HN, ST-NI, NI-HN, MR-ES, ES-FS, TR-PN, ML-MR, AD-VR, DA-LV, AD-RE, RE-VR, JL-KR, RT-ER, ER-SS, RT-BT, DM-GA, MR-CR, DM-FI, FI-MR; *Stars*, 2, FI, DM; *Chains*, 2, GA-DM-FI-MR-CR, SS-ER-RF-BT-SR; *Triangles*, 2, AD-VR-RE, HN-NI-ST; *Inter-sexual Attractions*, 1.

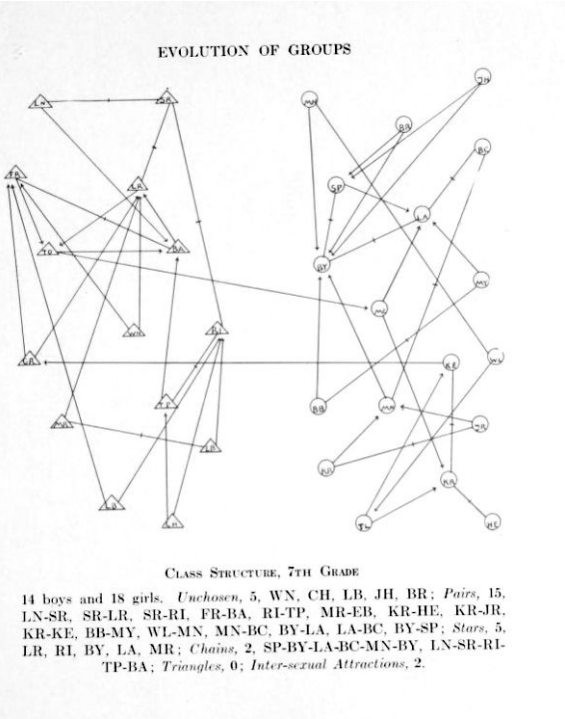


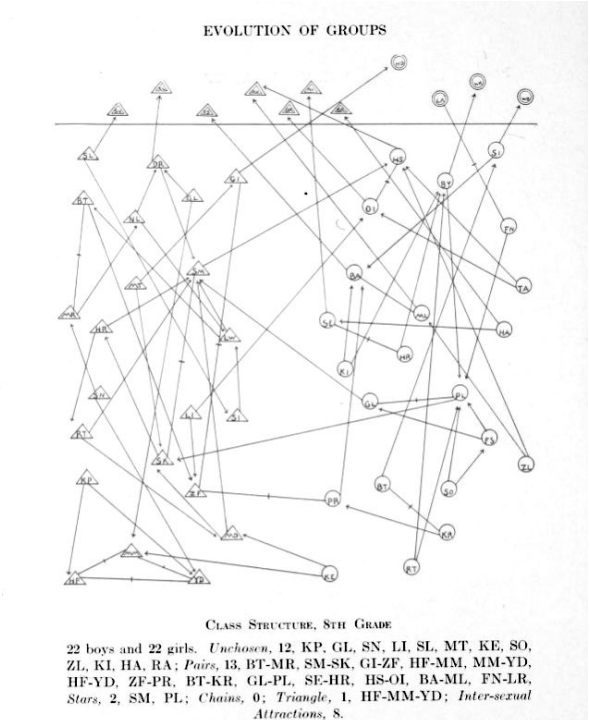
CLASS STRUCTURE, 6TH GRADE

18 boys and 21 girls. *Unchosen*, 6, FS, HD, ML, GA, BR, WL; *Pairs*, 26, WL-WT, WT-SH, GO-TI, GO-GI, TI-GI, PI-WI, PI-SH, NL-PI, YG-FR, YG-SH, SH-FR, DM-AL, SL-BY, SL-LP, LP-BY, FE-TU, TU-KR, NS-HI, HR-AR, OE-AR, OE-CL, CL-KR, KN-PT, WL-LY, CA-SH, KR-HD; *Stars*, 1, OE; *Chains*, 2, LY-WL-WT-SH-CA, HR-AR-OE-SL-KR;

Moreno J. Who Shall Survive: A New Approach to the Problem of Human Interrelations. Washington: Nervous and Mental Disease Publishing Co. 1934

Network analysis has a long tradition in social sciences





Moreno J. Who Shall Survive: A New Approach to the Problem of Human Interrelations. Washington: Nervous and Mental Disease Publishing Co. 1934

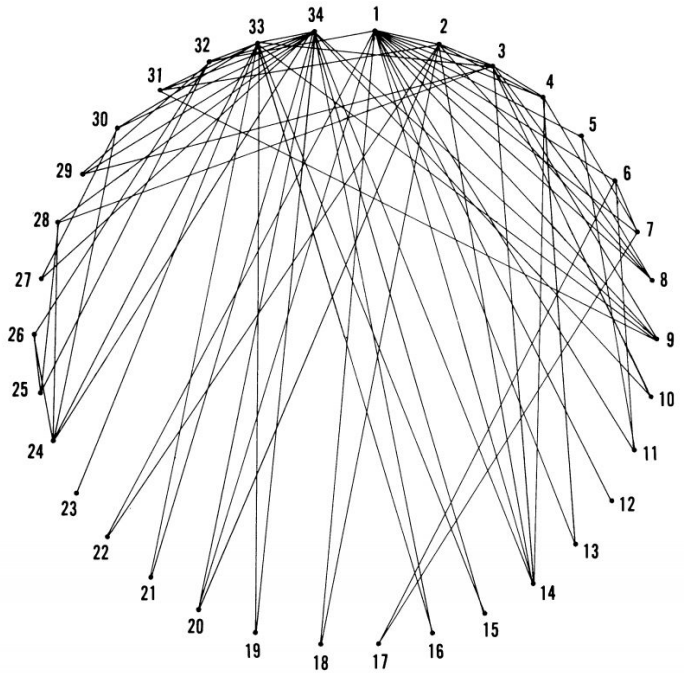
Karate Club (observation in 1972, paper in 1976)

MATRIX OF RELATIONSHIPS IN THE CLUB: THE MATRIX *E*

Individual Number

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
2	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
3	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
4	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
17	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
20	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
22	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 1
Social Network Model of Relationships in the Karate Club



Zachary, W. (1976). An Information Flow Model for Conflict and Fission in Small Groups1. Journal of Anthropological Research, 33. https://doi.org/10.1086/jar.33.4.3629752

‘The strength of weak ties’ by Mark Granovetter (1973)

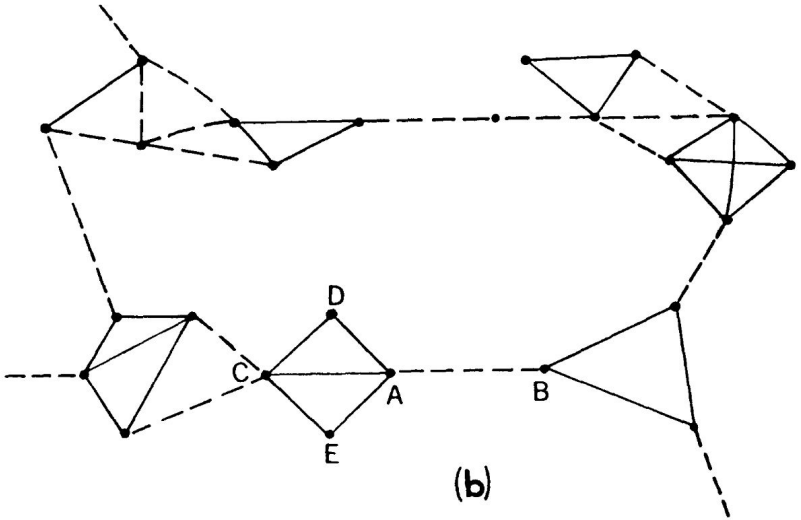


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

Granovetter, M. S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360–1380. http://www.jstor.org/stable/2776392

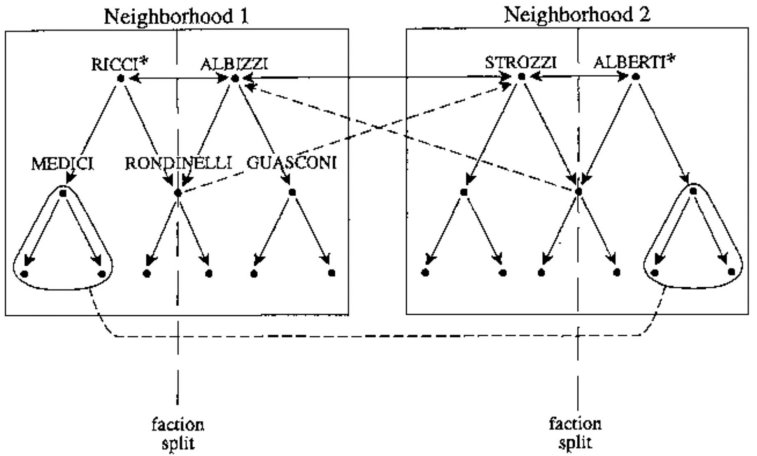
And in the 1990es and early 2000s social scientists started trying it on the objects of research typical of the humanities

Social scholars applying network analysis to historical sources in 1993:

DATA SOURCES AND SELECTION

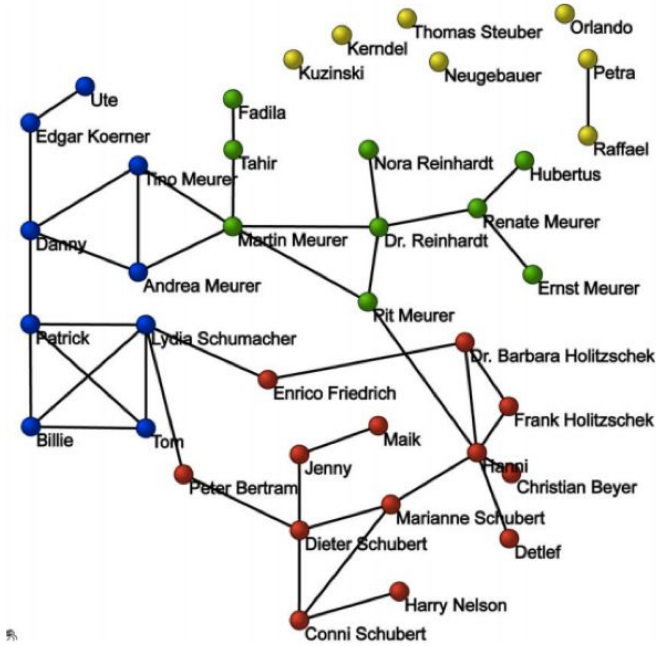
This article is empirically possible because of the thorough and impressive work of many historians of Florence. In particular, we build on the work of Dale Kent, whose book, *The Rise of the Medici* (1978), in the tradition of Lewis Namier (1929), is an intimate prosopographical description of the network foundations both of the Medici party, or faction, and of the looser alliance system of their opponents.[12] From the detailed text of this account, we coded a core network data set, which consists of information on the following nine types of relations among early 15th-century Florentine elite families: (*a*) one type of kinship relation—intermarriage ties,[13] (*b*) four types of economic relations—trading or business ties, joint ownerships or partnerships, bank employment, and real estate ties,[14] (*c*) two types of "political" relations—patronage and per-

American Journal of Sociology



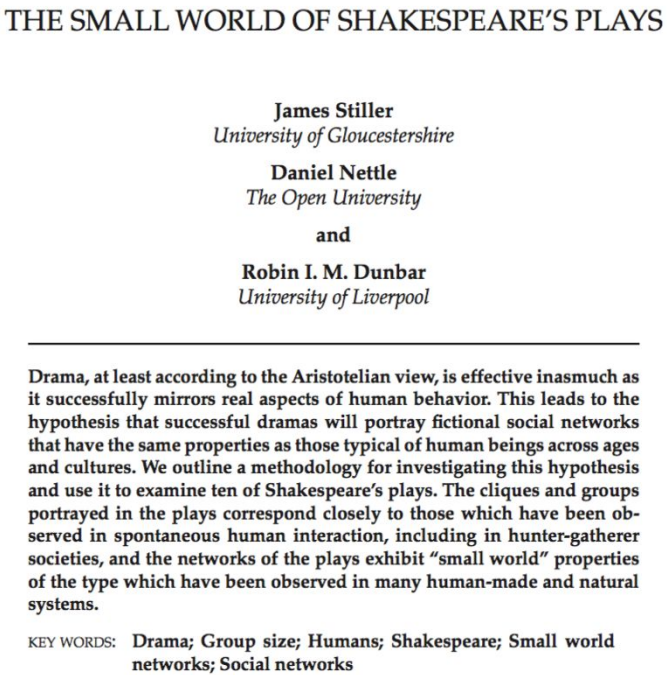
Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. American Journal of Sociology, 98(6), 1259–1319.

Social scholars applying network analysis to literary sources in 1998:



Schweizer T., Schnegg M. Die soziale Struktur der. „Simple Storys“: Eine Netzwerkanalyse. 1998

Social scholars applying network analysis to literary sources in 2003:



THE SMALL WORLD OF SHAKESPEARE'S PLAYS

James Stiller
University of Gloucestershire

Daniel Nettle
The Open University

and

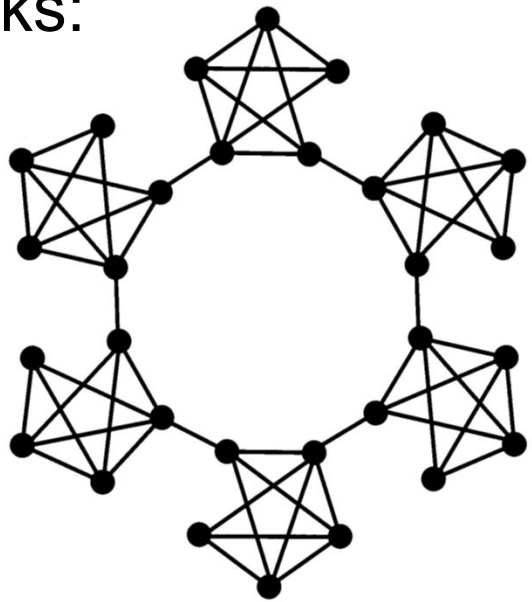
Robin I. M. Dunbar
University of Liverpool

Drama, at least according to the Aristotelian view, is effective inasmuch as it successfully mirrors real aspects of human behavior. This leads to the hypothesis that successful dramas will portray fictional social networks that have the same properties as those typical of human beings across ages and cultures. We outline a methodology for investigating this hypothesis and use it to examine ten of Shakespeare's plays. The cliques and groups portrayed in the plays correspond closely to those which have been observed in spontaneous human interaction, including in hunter-gatherer societies, and the networks of the plays exhibit "small world" properties of the type which have been observed in many human-made and natural systems.

KEY WORDS: **Drama; Group size; Humans; Shakespeare; Small world networks; Social networks**

James Stiller, Daniel Nettle, and Robin I. M. Dunbar (2003) The Small World of Shakespeare's Plays.Human Nature 14(4):397---408.

Small-world networks:

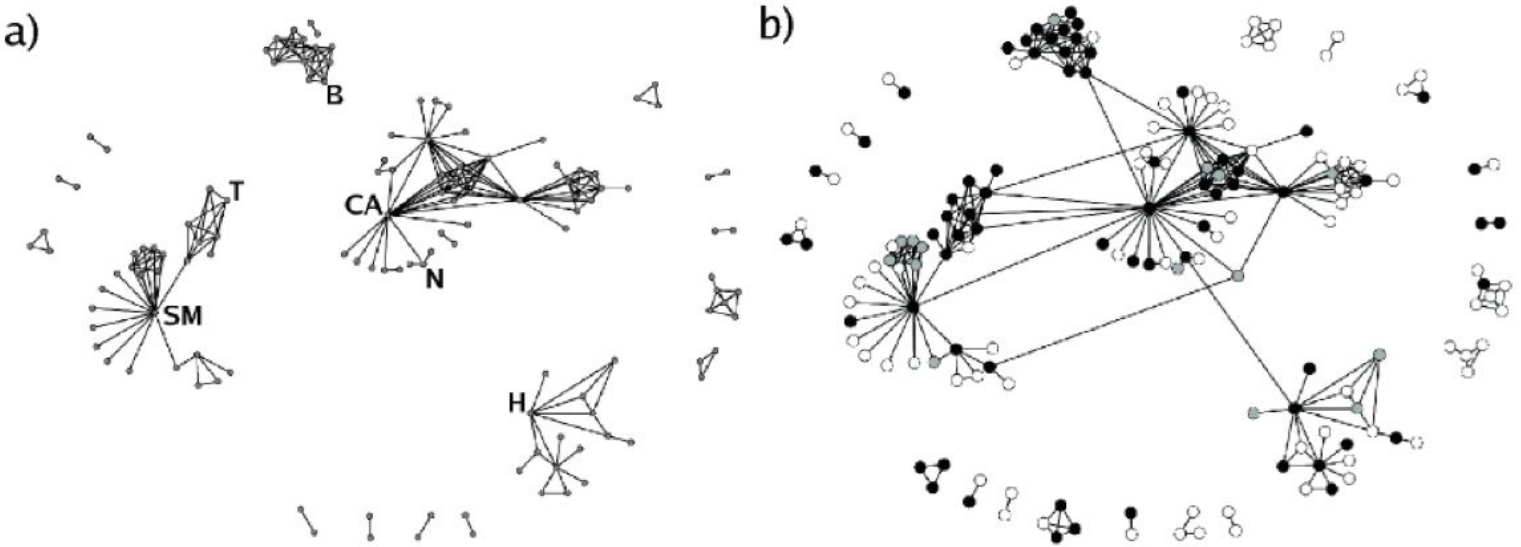


From Watts (1999)

“A small-world network is a mathematical graph in which most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other. Due to this, most neighboring nodes can be reached from every other node by a small number of hops or steps” (en.wikipedia.org/wiki/Small-world_network)

Early 2000-es: first large-scale research in the humanities:

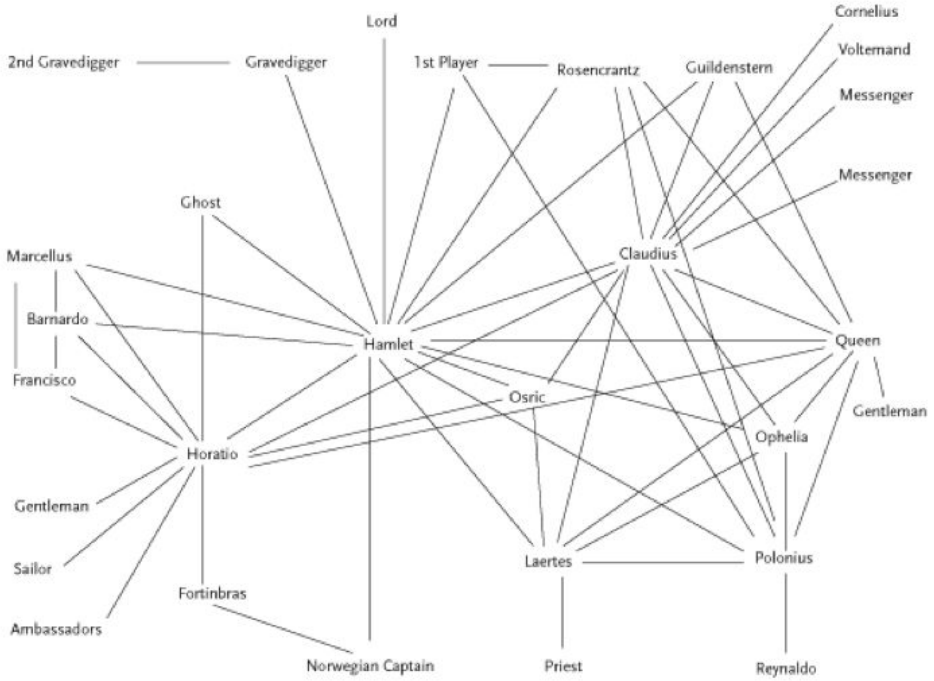
The Marvel Universe, networks extracted from 10k+ comic books:



1. Alberich, R., Miro-Julia, J., Rossello, F. (2002), Marvel universe looks almost like a real social network.
2. P. M. Gleiser. How to become a superhero. Journal of Statistical Mechanics: Theory and Experiment, (09):P09020, 2007. (picture source)

But the actual humanities scholars were missing

Literary network analysis reinvented by *literary* scholars like Franco Moretti in 2010-s



Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the “Distant Reading” book)

Moretti advocates the use of networks:

"[Networks] make visible specific ‘regions’ within the plot as a whole <...>. Take the characters who are connected to both Claudius and Hamlet <...>: except for Osric and Horatio, whose link to Claudius is however extremely tenuous, they are all killed <...>; what is truly deadly, is the characters’ position in the network, chained to the warring poles of king and prince. Outside of that bold region, no one dies in Hamlet. The tragedy is all there".

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the “Distant Reading” book)

Moretti advocates the use of networks:

"[Networks] make visible specific ‘regions’ within the plot as a whole <...>. Take the characters who are connected to both Claudius and Hamlet <...>: except for Osric and Horatio, whose link to Claudius is however extremely tenuous, they are all killed <...>; what is truly deadly, is the characters’ position in the network, chained to the warring poles of king and prince. Outside of that bold region, no one dies in Hamlet. The tragedy is all there".

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the “Distant Reading” book)

Moretti advocates the use of networks:

"[Networks] make visible specific ‘regions’ within the plot as a whole <...>. Take the characters who are connected to both Claudius and Hamlet <...>: except for Osric and Horatio, whose link to Claudius is however extremely tenuous, they are all killed <...>; what is truly deadly, is the characters’ position in the network, chained to the warring poles of king and prince. Outside of that bold region, no one dies in Hamlet. The tragedy is all there".

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the “Distant Reading” book)

Network as a reduced model for plot that makes certain key features of a play more visible

Moretti comparing networks to an X-ray:

"You reduce the text to characters and interactions, abstract them from everything else, and this process of reduction and abstraction makes the model obviously much less than the original object — just think of this: I am discussing Hamlet, and saying nothing about Shakespeare's words — but also, in another sense, much more than it, because a model allows you to see the underlying structures of a complex object. It's like an X-ray: suddenly, you see the region of death <...> which is otherwise hidden by the very richness of the play".

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the "Distant Reading" book)

Moretti comparing networks to an X-ray:

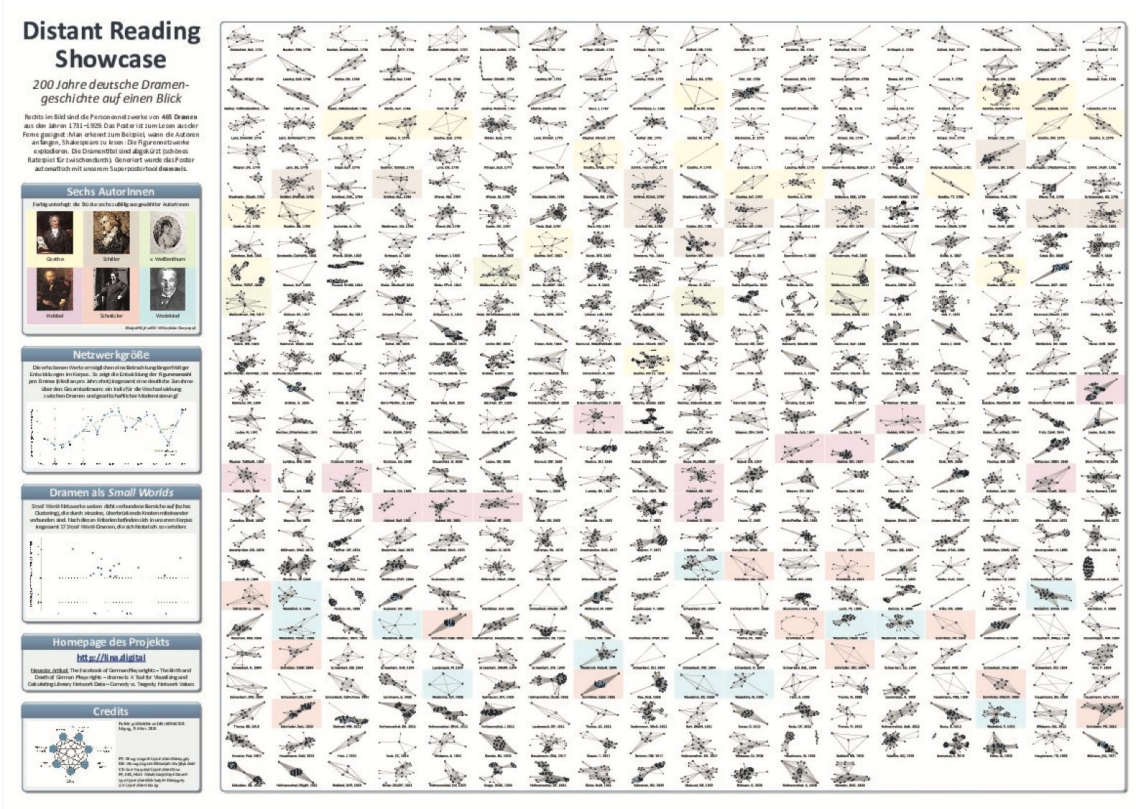
"You reduce the text to characters and interactions, abstract them from everything else, and this process of reduction and abstraction makes the model obviously much less than the original object — just think of this: I am discussing Hamlet, and saying nothing about Shakespeare's words — but also, in another sense, much more than it, because a model allows you to see the underlying structures of a complex object. It's like an X-ray: suddenly, you see the region of death <...> which is otherwise hidden by the very richness of the play".

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the "Distant Reading" book)

And today literary network analysis is

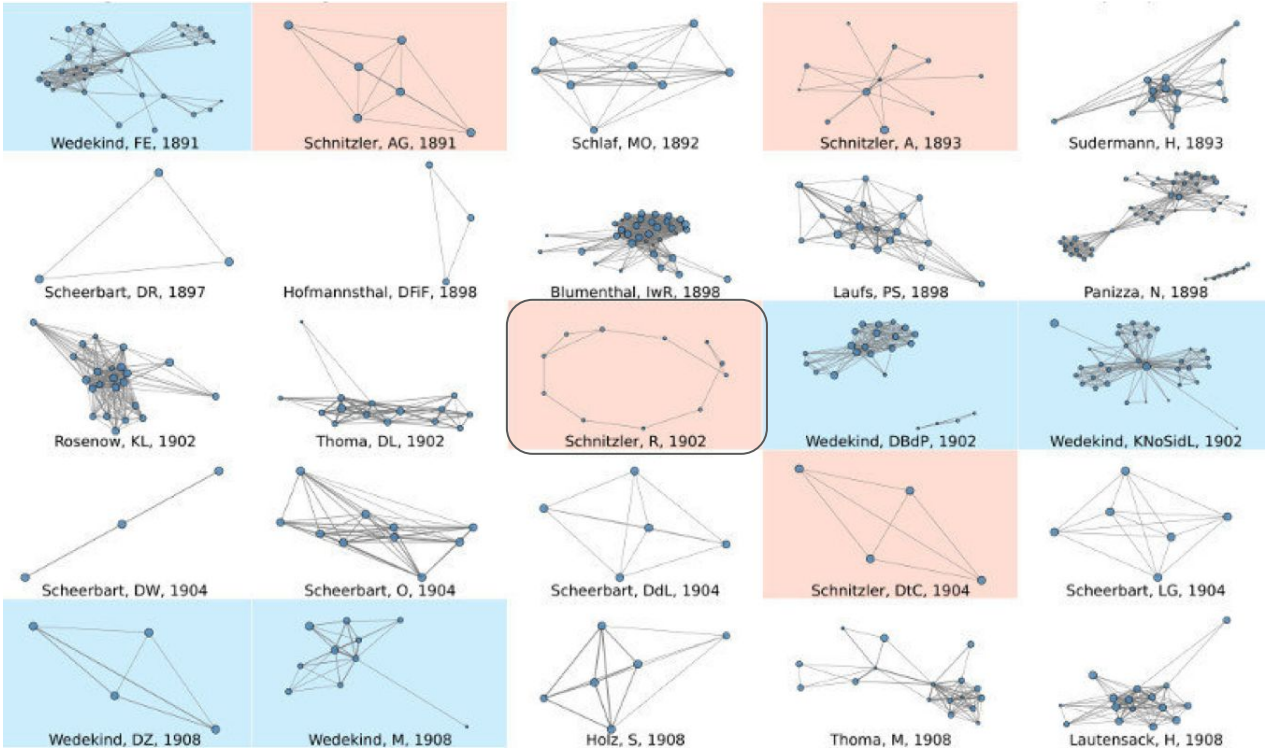
a big thing

Network analysis of Drama on Dracor data (large-scale)

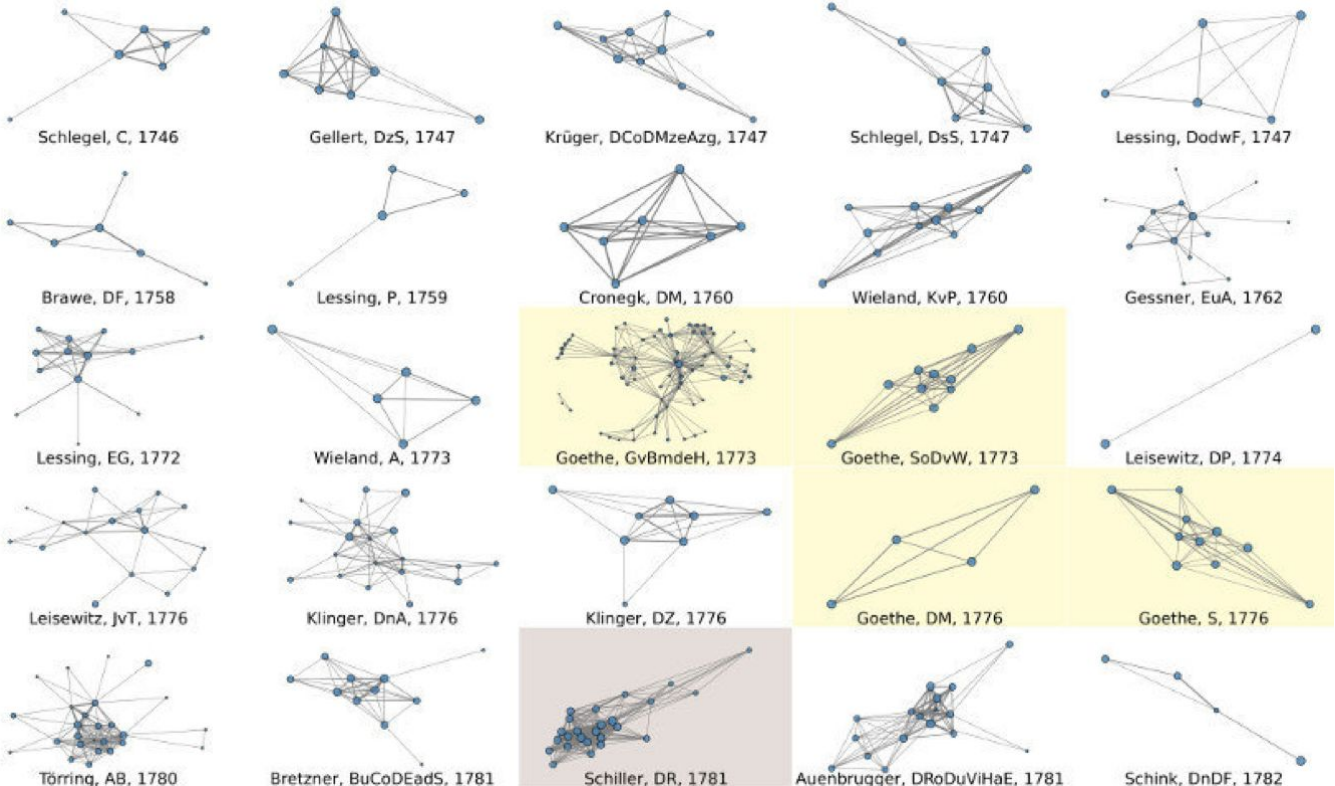


Character networks of 465 plays at a glance. Presented at DHd2016. DOI: 10.6084/m9.gshare.3101203.v1.

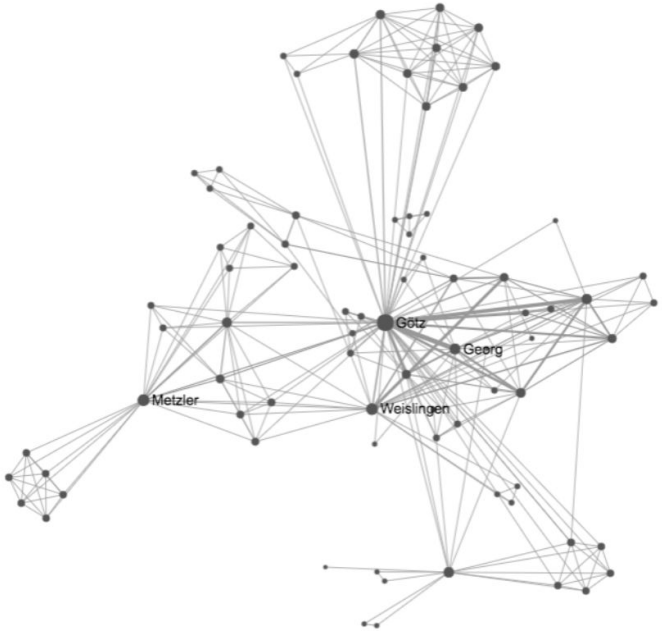
Bird’s eye view allows you to detect interesting structures (even before we start the *formal* analysis):



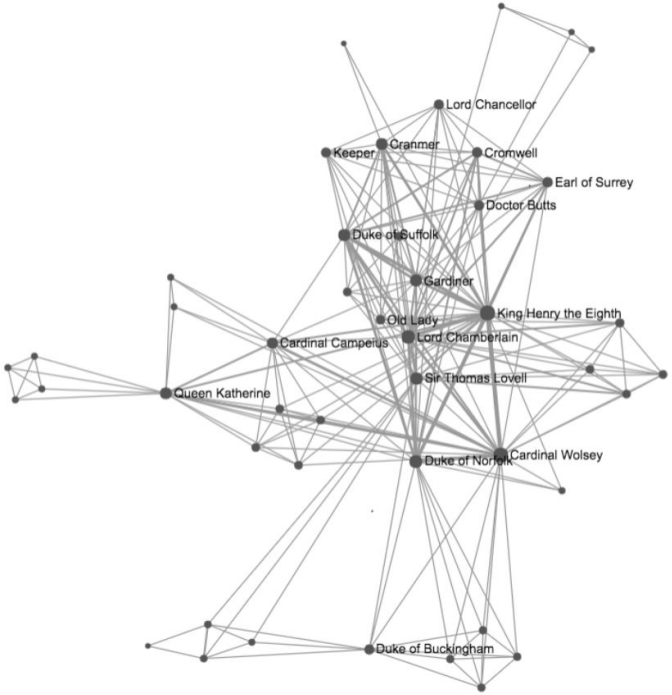
Bird’s eye view allows you to detect interesting structures (even before we start the *formal* analysis):



Goethe-Shakespeare structural parallelism:

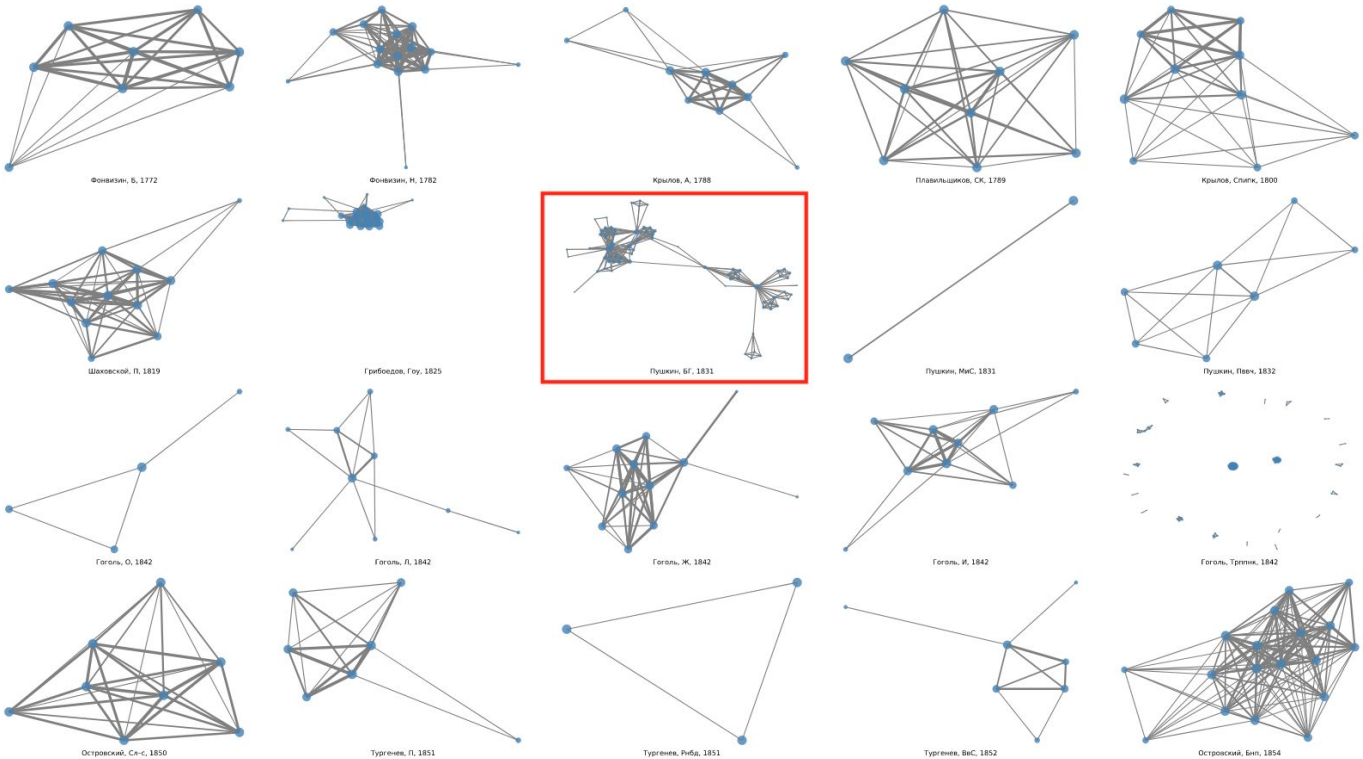


‘Historical’ Goethe (Götz von Berlichingen)

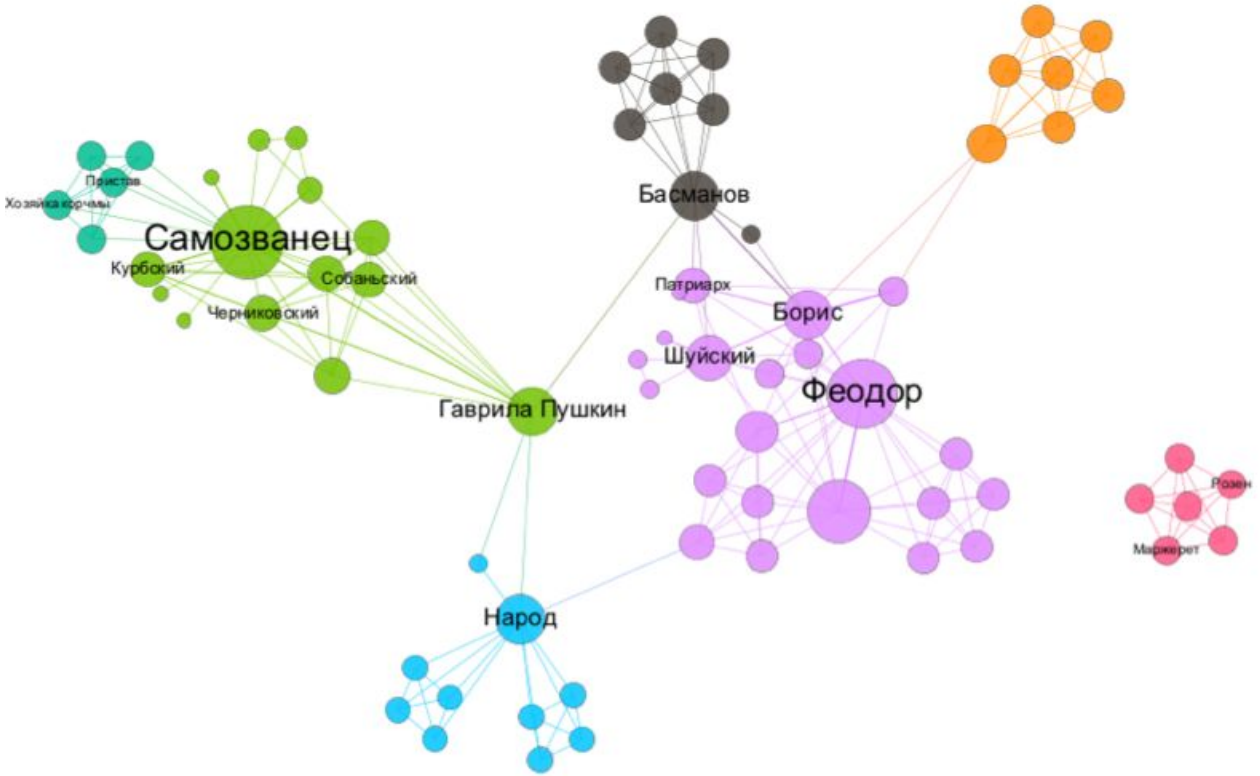


‘Historical’ Shakespeare (Henry VIII)

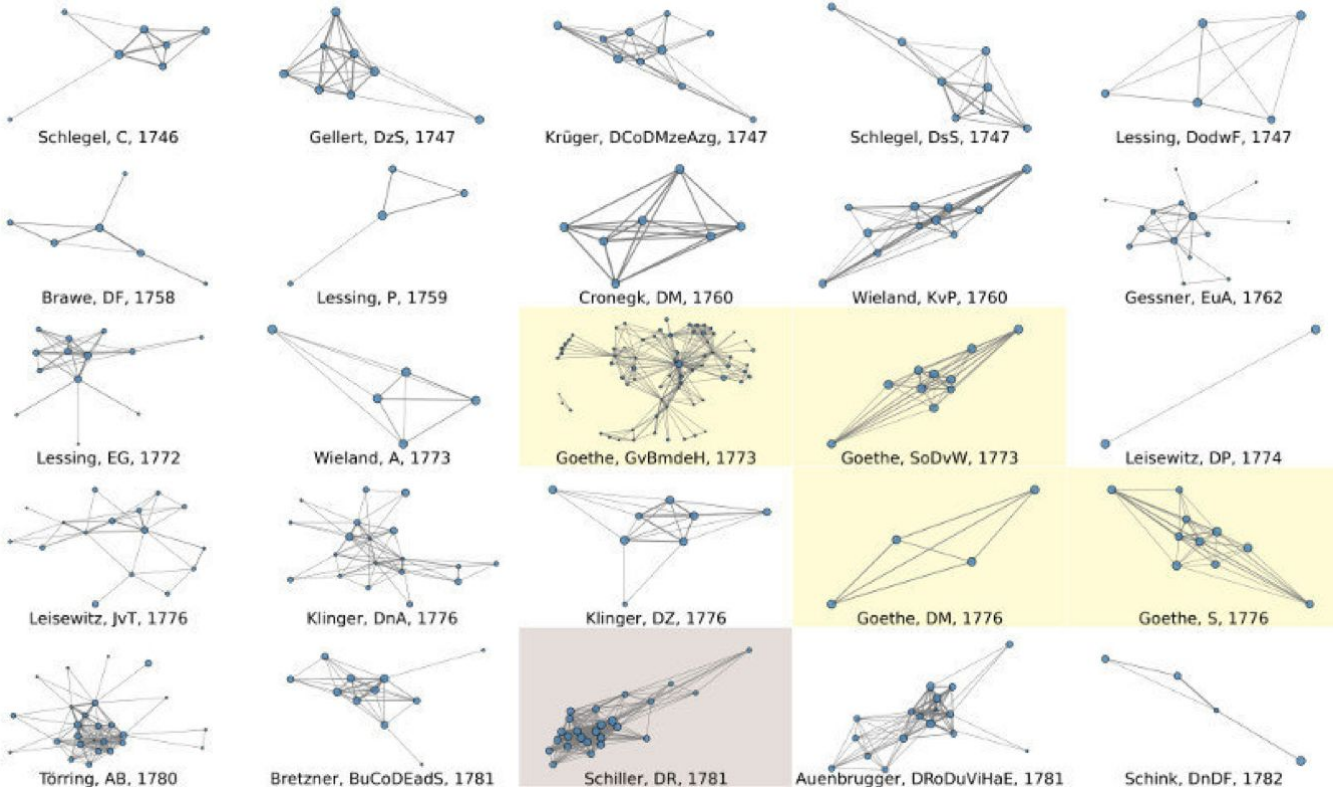
This process can be traced in other literatures (e.g. Ru)



Zooming in on Boris Godunov by A.S. Pushkin:



Same as Goethe: too much historical Shakespeare…



"Не смущаемый никаким иным влиянием, Шекспиру я подражал в его вольном и широком изображении характеров, в небрежном и простом составлении планов…"

А. С. Пушкин. Собрание сочинений в десяти томах. Том 6. Критика и публицистика, с. 300

("Unaffected by any other [literary] influence, I was following Shakespeare in his free and broad depiction of characters, in his careless and simple construction of settings...")

A.S. Pushkin, my translation

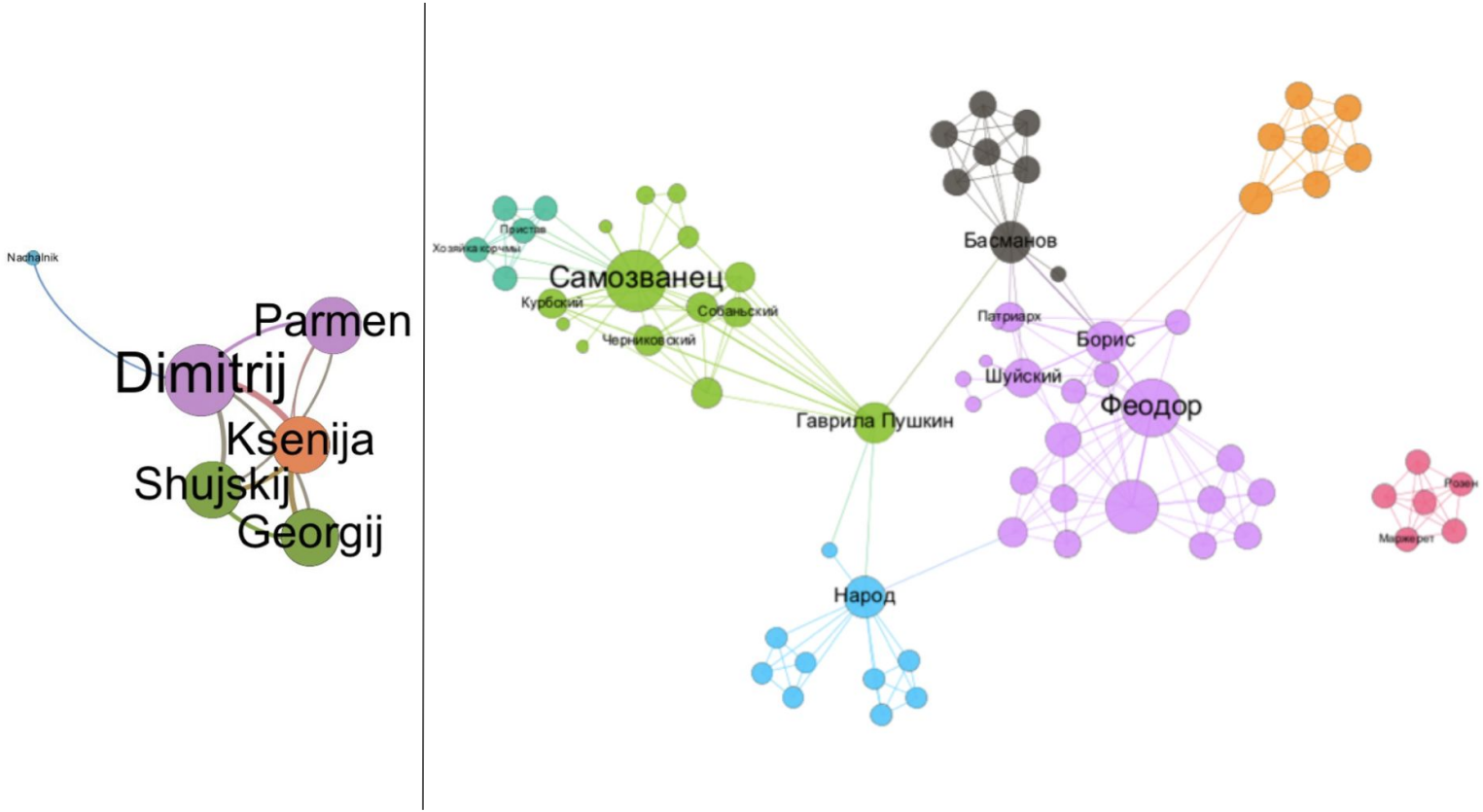
"Твердо уверенный, что устарелые формы нашего театра требуют преобразования, я расположил свою трагедию по системе Отца нашего Шекспира, и принес ему в жертву пред его алтарь два классические единства, и едва сохранил последнее".

А. С. Пушкин. Собрание сочинений в десяти томах. Том 6. Критика и публицистика, с. 250

"(Being quite confident that the obsolete forms of our theatre require reformation, I laid out my tragedy according to the system of our Father Shakespeare, and sacrificed two of the three classical unities, and barely managed to preserve the third").

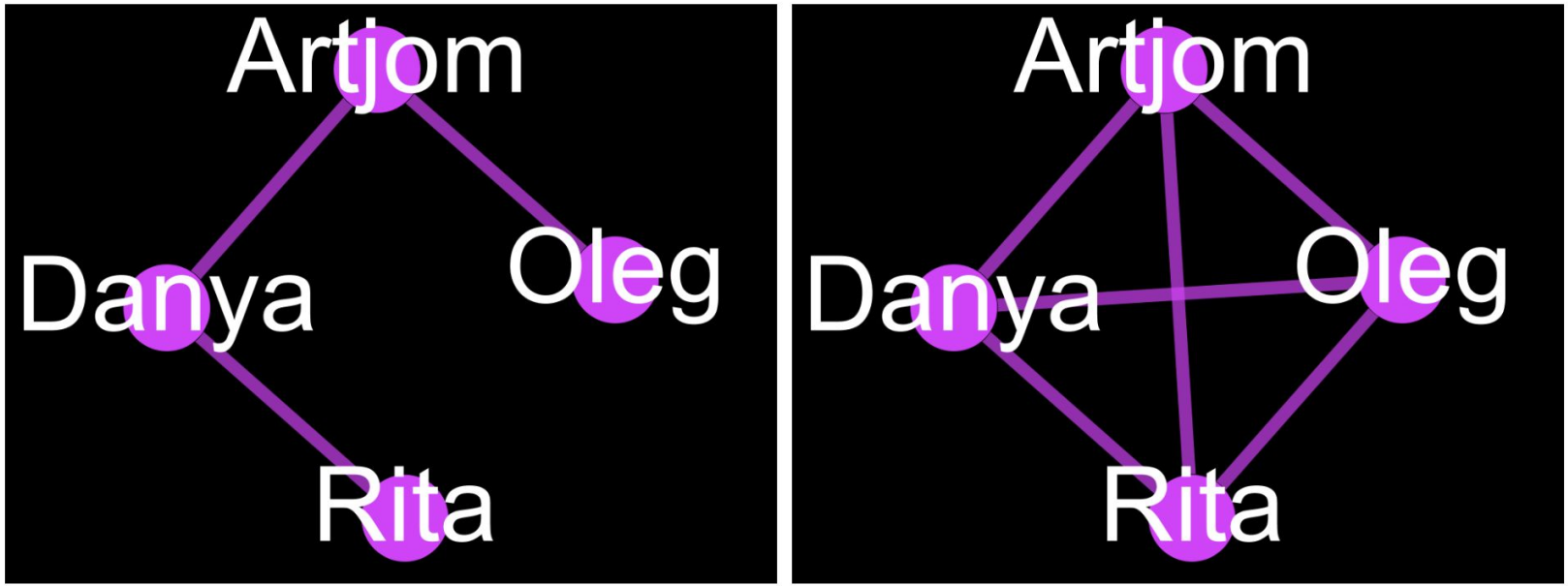
A.S. Pushkin, my translation

a Classicist vs a 'Historical Shakespearean’ play:



Can we make it more formal?

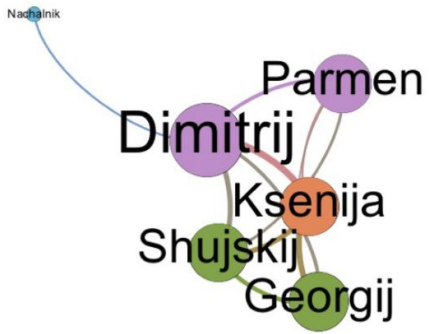
Density of a network:



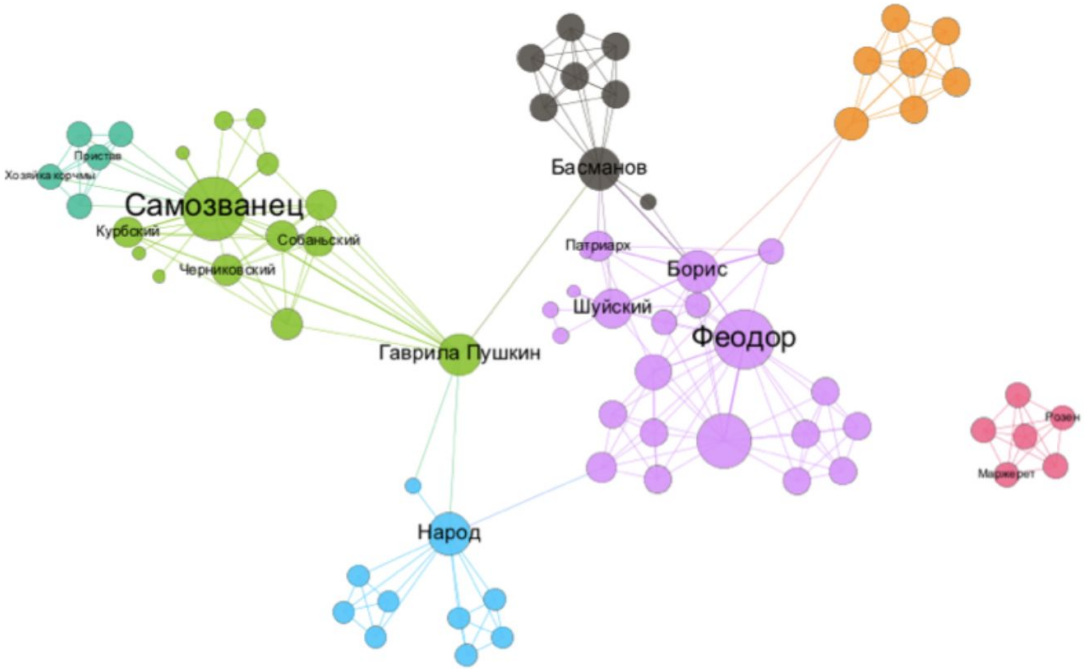
Density = 0.5 (3/6) Density = 1 (6/6)

a Classicist vs a 'Historical Shakespearean' play:

Density: **0.73**



Density: **0.11**



Feel free to double-check yourself on DraCor.org

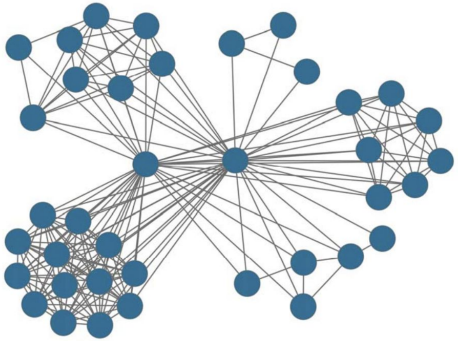
https://dracor.org/rus/sumarokov-dimitrij-samozvanets

https://dracor.org/rus/pushkin-boris-godunov

Zooming even deeper to the level of individual characters/actors

Network centralities can be used in CLS as a feature to differentiate characters:

Frank Wedekind: »Franziska« (1912)



Character	Count Ranking	Network Ranking
Franziska	1	1
Veit Kunz	2	2
Herzog	15	3
Laurus Bein	3	16
Gespensterschreck	3	18
Breitenbach	16	5
Gislind	18	6
Sophie	20	4
Spreizfüsschen	3	23
Mausi	3	26
Herzogin	17	13
Rohrdommel	3	27
Fahrstuhl	24	9
...	...	...

Fischer F., Trilcke P., Kittel C., Milling C., Skorinkin D. To Catch a Protagonist: Quantitative Dominance Relations in German Language Drama (1730–1930), in: Digital Humanities 2018: Book of Abstracts / Libro de resúmenes. Mexico : Red de Humanidades Digitales A. C., 2018. P. 193-201.

Let's get back to Moretti for a sec:

Though Horatio is an old fixation of mine, I had never fully understood his role in Hamlet until I looked at the play's network structure.

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the "Distant Reading" book)

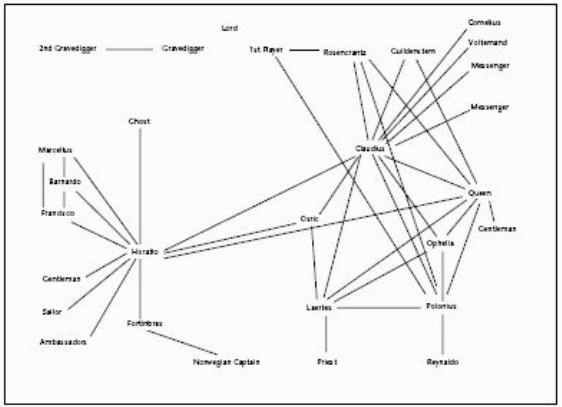


Figure 7: Hamlet without Hamlet

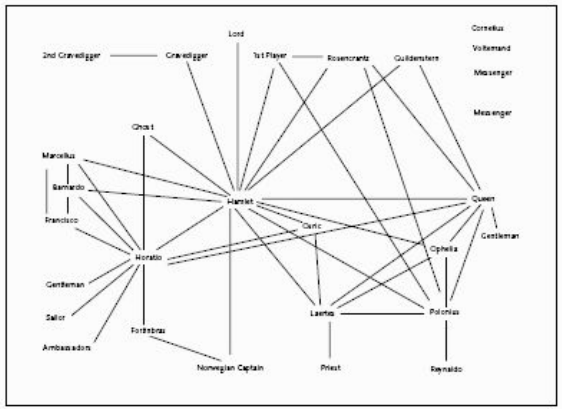


Figure 8: Hamlet without Claudius

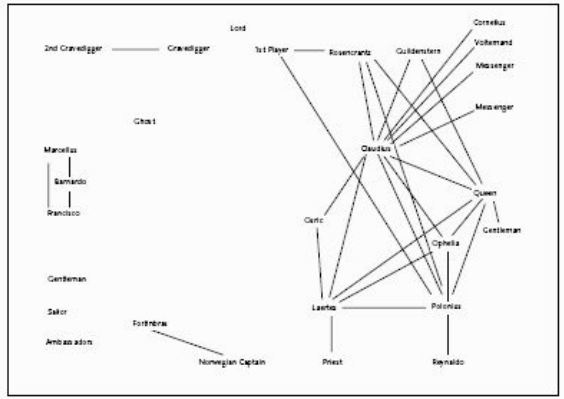


Figure 9: Hamlet without Hamlet and Horatio

Why is the protagonist significant here? Not for what is 'in' it; not for its essence, but for its function in the stability of the network. And stability has clearly much to do with centrality, but is not identical to it. Take the second-most-central character of the play: Claudius. In quantitative terms, Claudius is almost as central as Hamlet (average distance of 1.62, versus 1.45); but in structural terms not so, when we remove him from the network (Figure 8), what happens is that a handful of peripheral characters are affected, but the network as a whole not much. Even if we remove, first Hamlet, and then Claudius, his subtraction does not do much. But if we remove, first Hamlet, and then *Horatio* (Figure 9), then the fragmentation is so radical that the Ghost and Fortinbras—which is to say, the beginning and the ending of the play—are completely severed from each other and from the rest of the plot. *Hamlet* no longer exists. And yet, Horatio is slightly less central than Claudius in quantitative terms (1.69 versus 1.62). Why is he so much more important in structural terms?

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the "Distant Reading" book)

Moretti 'dissecting' the network:

<..> if we remove, first Hamlet, and then Horatio <..>, then the fragmentation is so radical that the Ghost and Fortinbras — which is to say, the beginning and the ending of the play — are completely severed from each other and from the rest of the plot. Hamlet no longer exists. And yet, Horatio is slightly less central than Claudius in quantitative terms

Franco Moretti. Network Theory, Plot Analysis (Chapter 9 of the "Distant Reading" book)

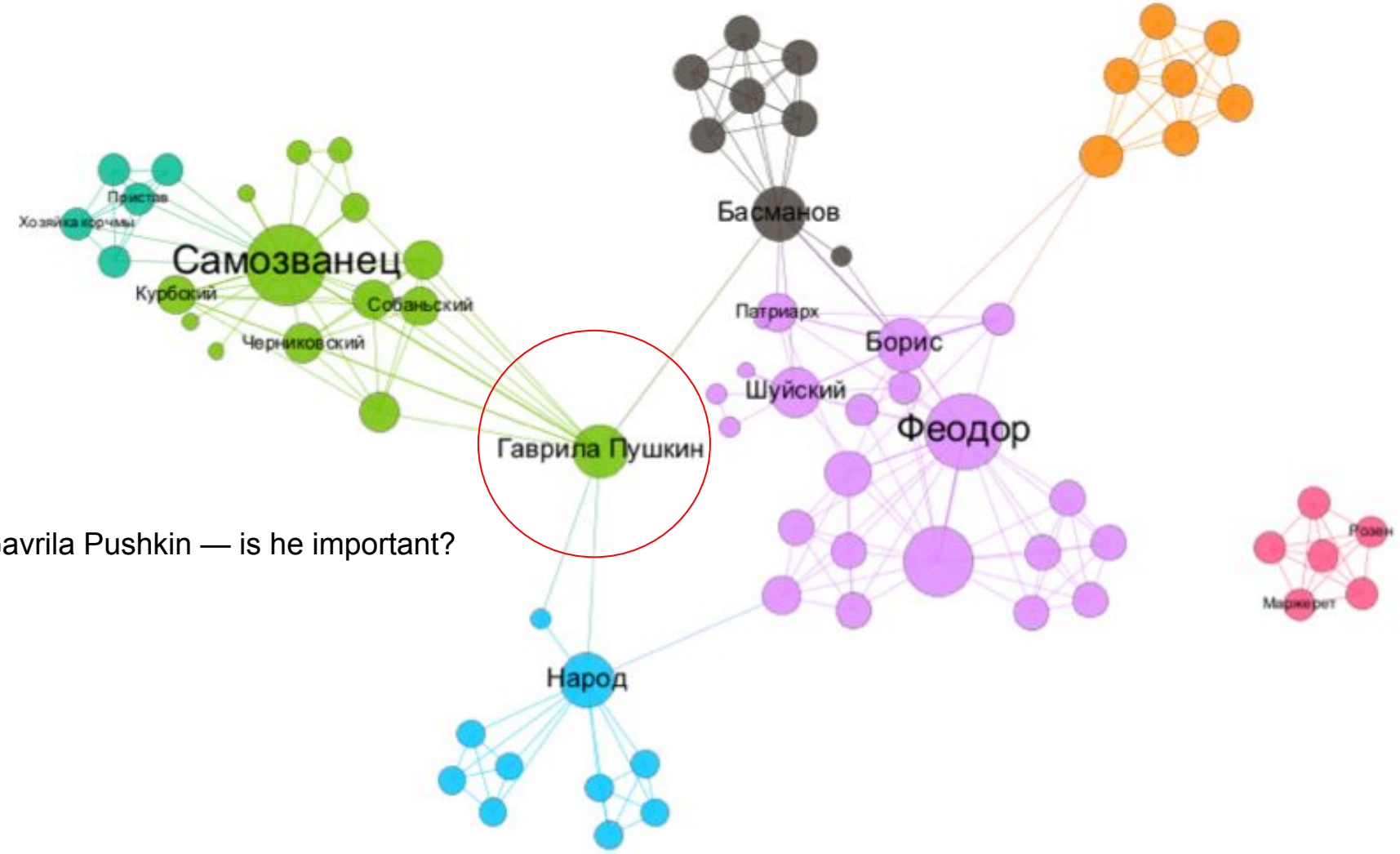
Makes sense to comapre different network centralities here

Checking Moretti's assumptions

Character	Degree	Betweenness
Hamlet	19	219,4
Claudius	16	131,6
Horatio	14	176,7

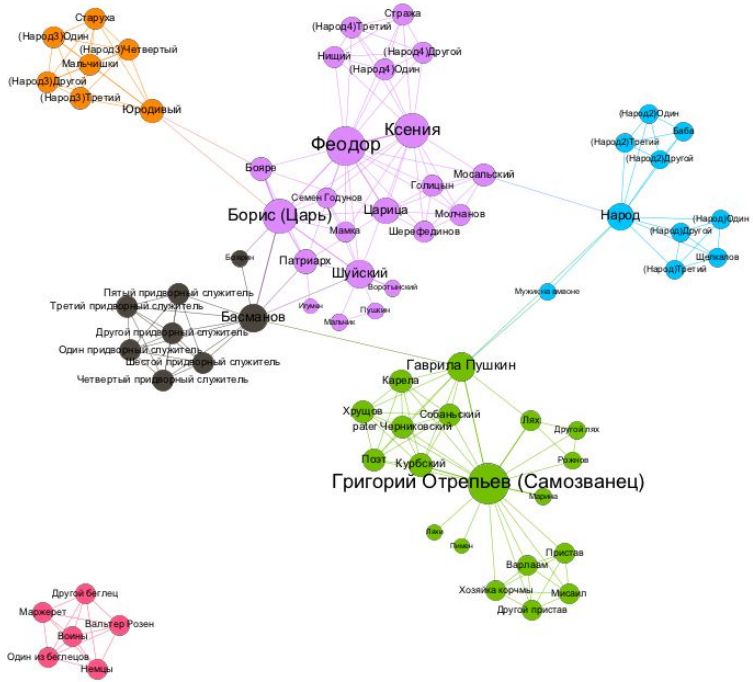
So, Moretti seems to have been right, even though he almost did not use the maths behind the network science. Horatio is less central in terms of degree (the most basic centrality), but he is more important for holding the entire play together, which manifests itself in higher betweenness centrality

And now back to Pushkin(s)

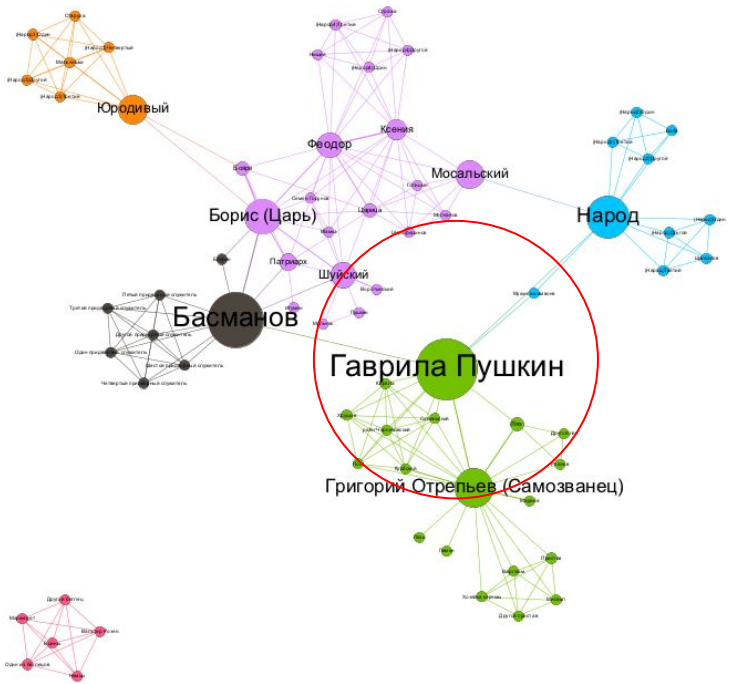


Gavrila Pushkin — is he important?

Yes, he kind of is: moderate degree but top betweenness



Node&label size = Node Degree



Node&label size = Betweenness Centrality

A.S. Pushkin, 'My pedigree' (something we would typically learn at school):

<...>"About the Tsars the Pushkins moved"<...>

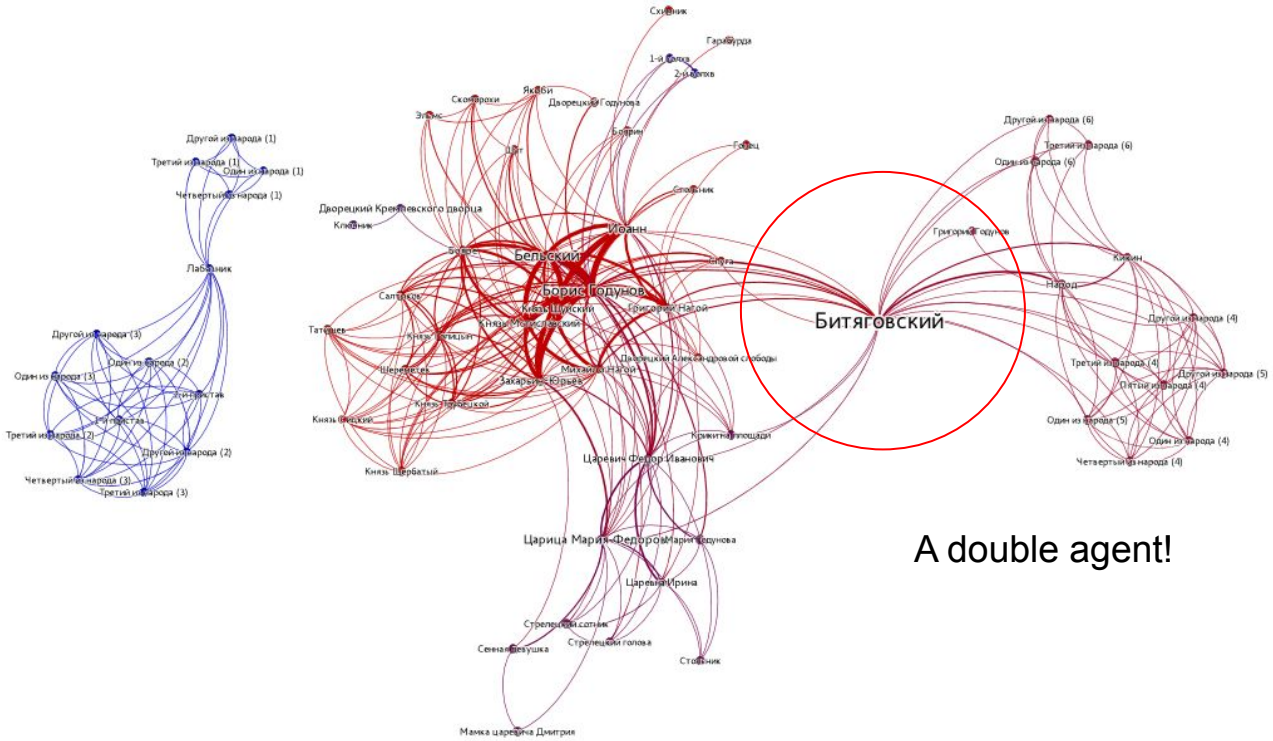
When the Romanoffs upon the throne

The nation called by its Chart --

We upon it laid our hands

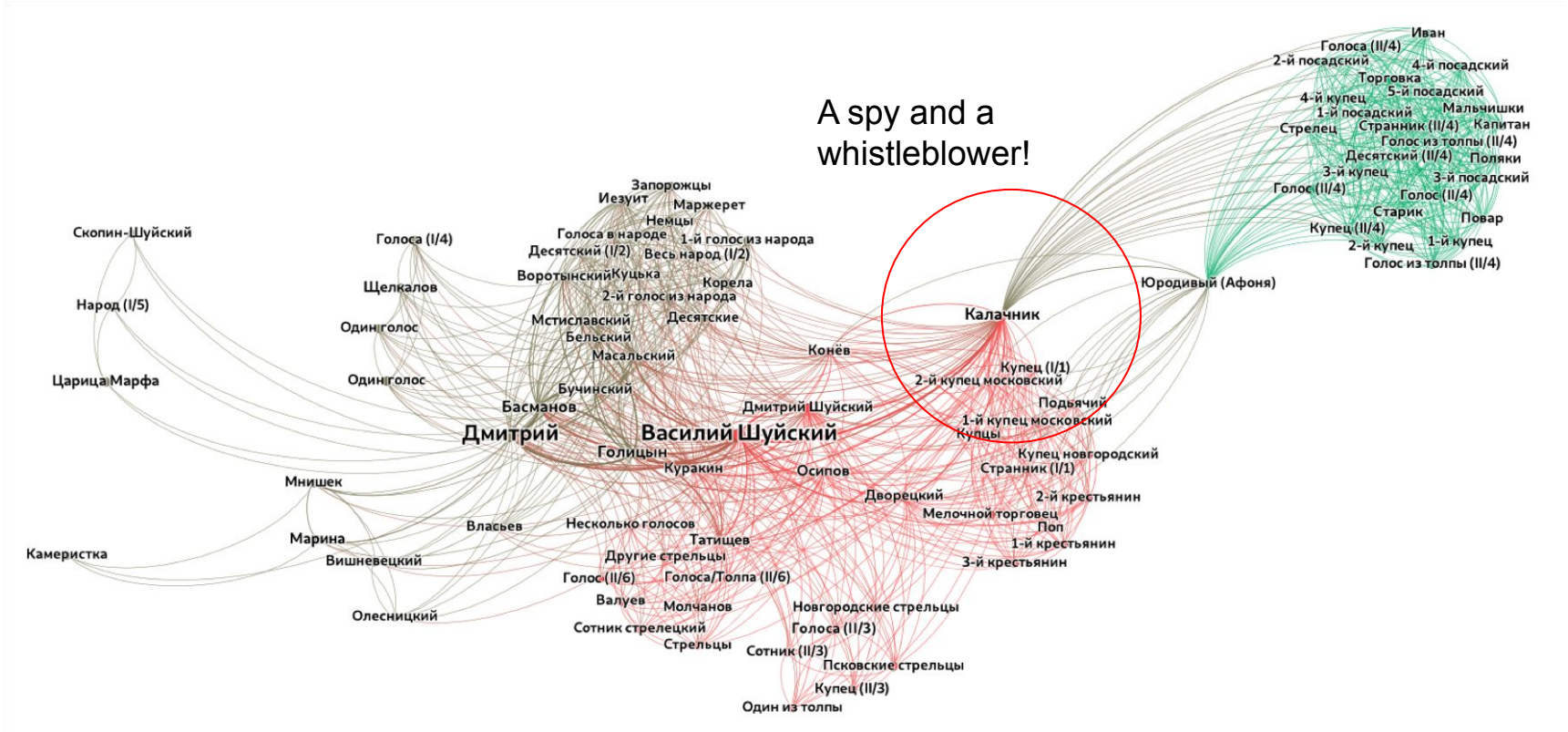
And now that we have a *formalization* of this type of character..
..and can look for them in more plays!

Other high-betweenness characters: Bityagovskiy in «The Death of Ivan The Terrible» by A.K. Tolstoy:



A double agent!

Kalachnik in «Dimitry the Imposter» by A.N. Ostrovsky



Some confirmation for ‘betweenness-bases spy tracking’ from historians:

Metadata, Surveillance and the Tudor State

by Ruth Ahnert and Sebastian E. Ahnert

‘We kill people based on metadata.’
General Michael Hayden, former director of the NSA, 2014.[1]

On 5 June 2013 the *Guardian* newspaper published an exclusive article, based on information leaked to them by Edward Snowden, that revealed a large-scale effort by the United States National Security Agency to collect domestic email and telephone metadata from the US telecommunications company Verizon. In the following days it would emerge that this was part of a more widespread and systematic programme, named Prism, which allowed the Agency to gain access from nine Internet companies to a wide range of digital information on foreign targets operating outside the United States. The NSA defended itself against allegations that its activities contravened the Fourth Amendment by arguing the difference between data and metadata: that the contents of the phone calls, emails and other communications

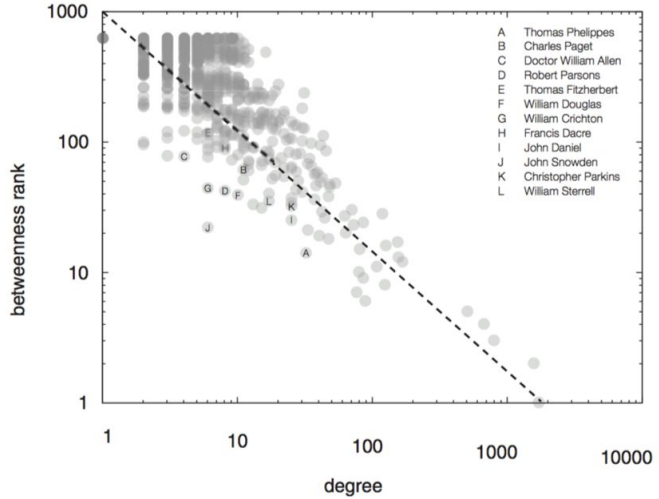


Fig. 4. This graph charts degree against the betweenness ranking of each person who corresponded in the 1590s.

Ahnert R., Ahnert S.E. Metadata, Surveillance and the Tudor State // History Workshop Journal. 2019. T. 87. C. 27–51.

Some confirmation for ‘betweenness-bases spy tracking’ from historians:

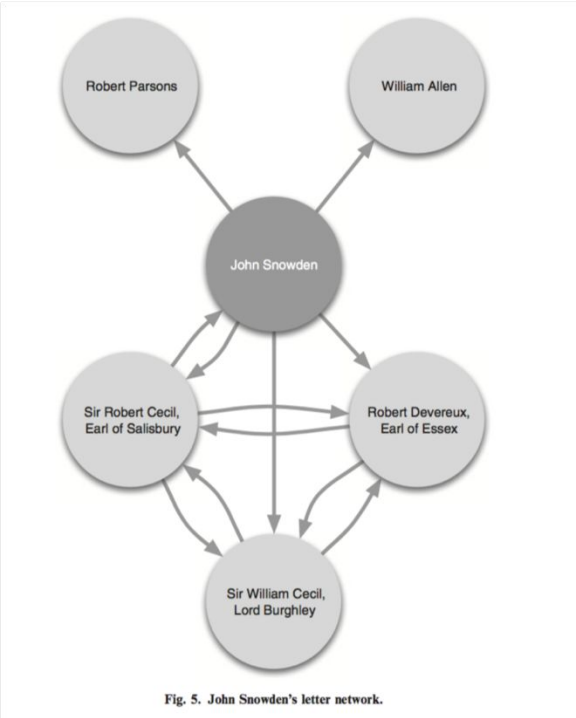


Fig. 5. John Snowden's letter network.

Ahnert R., Ahnert S.E. Metadata, Surveillance and the Tudor State // History Workshop Journal. 2019. T. 87. C. 27–51.

To sum up this long rant about literary network analysis:

- There are at least **two ways** to do network analysis of our research objects:
  - **Large(r)-scale**, when you have **multiple networks** and investigate their high-level properties (structure, density, small-world-ness)
  - **Small(er)-scale**, when you dig into an **individual work** and **compare actors/characters** within one network or character space (e.g. with centralities)
- Both ways provide you with **formalizations** and **numbers**:
  - **Structural metrics** (density, small-worldness, clustering coefficient, diameter, average path length etc.) in case of large-scale (corpus-level) research
  - **Centralities** (degree, betweenness, closeness) and clustering of a network in case of small-scale (individual text level) research

Pause for discussion

Part 2: Technicalities of network data preparation

1. How do we store networks in machine-readable form?
2. How to represent your object of research as network?
3. Automation of network extraction

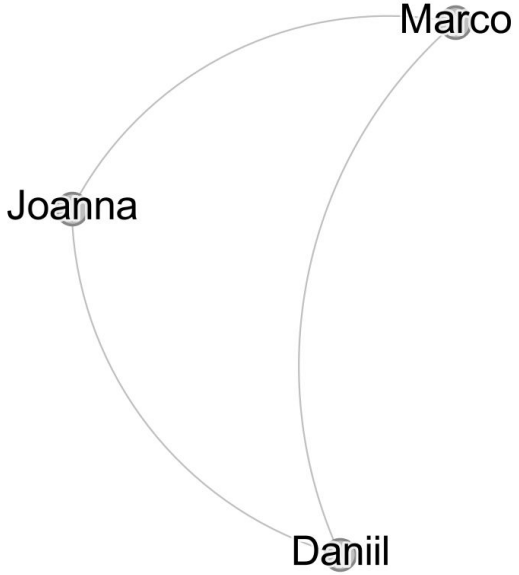
How do we store networks in machine-readable form?

a CSV table with network edges is a popular option:

Daniil,Joanna

Joanna,Marco

Daniil,Marco

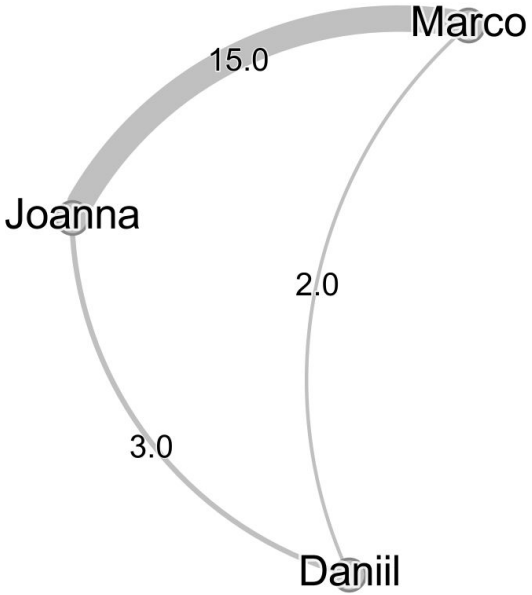


Adding weight information:

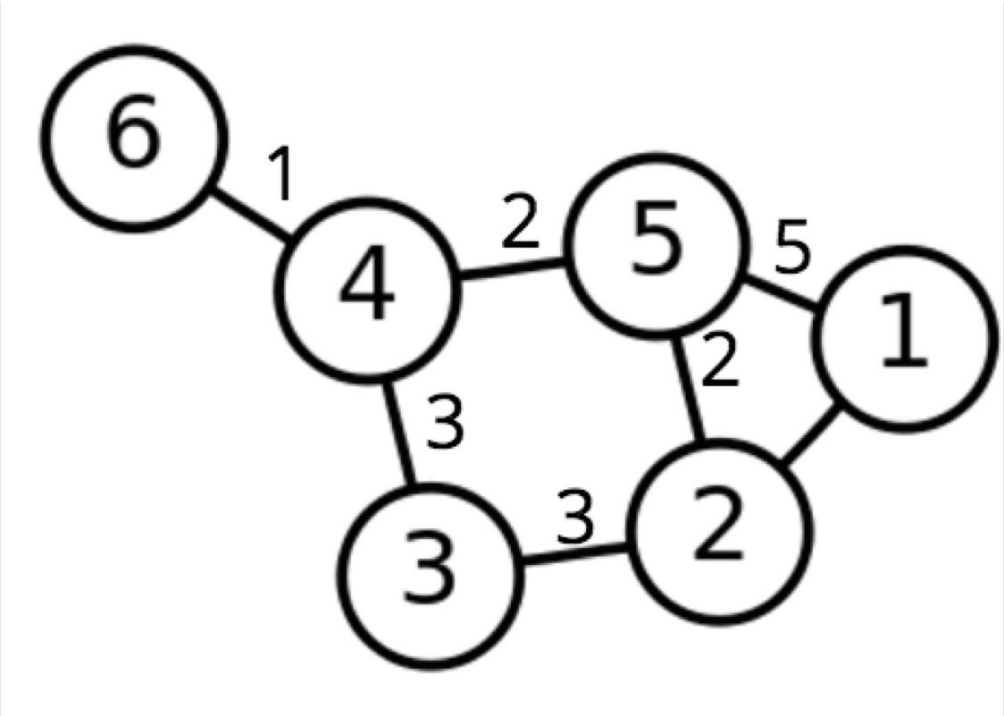
Daniil,Joanna,3

Joanna,Marco,15

Daniil,Marco,2



BTW we just introduced a weighted network (graph)

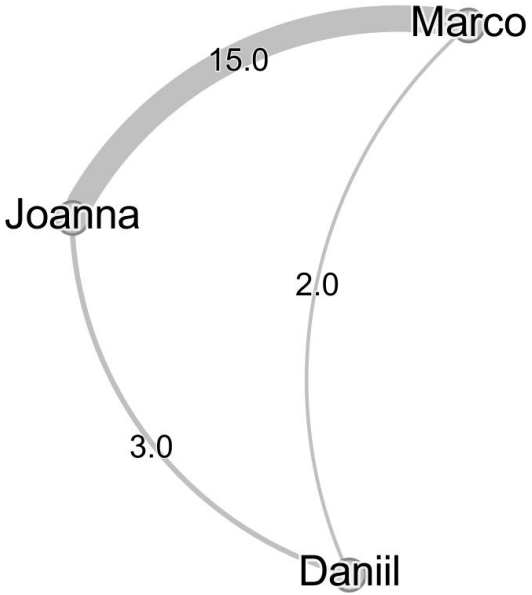


Adding weight information:

Daniil,Joanna,3

Joanna,Marco,15

Daniil,Marco,2



Another format: GraphML

XML-based format for network storage

```
<graphml>
  <graph id="G" edgedefault="undirected">
    <node id="n0"/>
    <node id="n1"/>
    <edge id="e1" source="n0" target="n1"/>
  </graph>
</graphml>
```

Another format: GEXF

Graph Exchange XML Format is another XML-based format, but this one is more complex and feature-rich. It allows for storing lots of metadata and visualisation parameters

```
<gexf>
    <meta lastmodifieddate="2018-03-20">
        <creator>Gexf.net</creator>
        <description>A hello world! file</description>
    </meta>
    <graph mode="static" defaultedgetype="directed">
        <nodes>
            <node id="0" label="Hello" />
            <node id="1" label="Word" />
        </nodes>
        <edges>
            <edge id="0" source="0" target="1" />
        </edges>
    </graph>
</gexf>
```

But a CSV table with network edges is the simplest

Daniil,Joanna,3

Joanna,Marco,15

Daniil,Marco,2

But a CSV table with network edges is the simplest. Here is what the CSV format accepted by Gephi looks like:

Source,Target,Weight

Daniil,Joanna,3

Joanna,Marco,15

Daniil,Marco,2

How do I obtain/produce such data?

There are graph datasets available in ready-to-use form:

- github.com/gephi/gephi/wiki/Datasets (datasets provided by the creators of Gephi)
- networks.skewed.de/ (a large repository of various networks — from dolphin friendships to Bitcoin transactions and scientific co-authorship)
- dracor.org (4330+ networks of theatre plays in 17 languages)

If you want to model your data as network on your own:

- First ask yourself, what is it you want to operationalize with a network
- Formalize the network for yourself:
  - What is a **node** (vertex) in our case?
  - What is an **edge** (connection)?
- Encode the edges in a machine-readable form — in other words, extract the network from a source (which can be text, historical objects, other types of sources, observations of life etc.)

Once you have a formalisation in your mind… there are several ways to actually extract the network

Many researchers build pipelines for automatically extracting networks from unstructured data

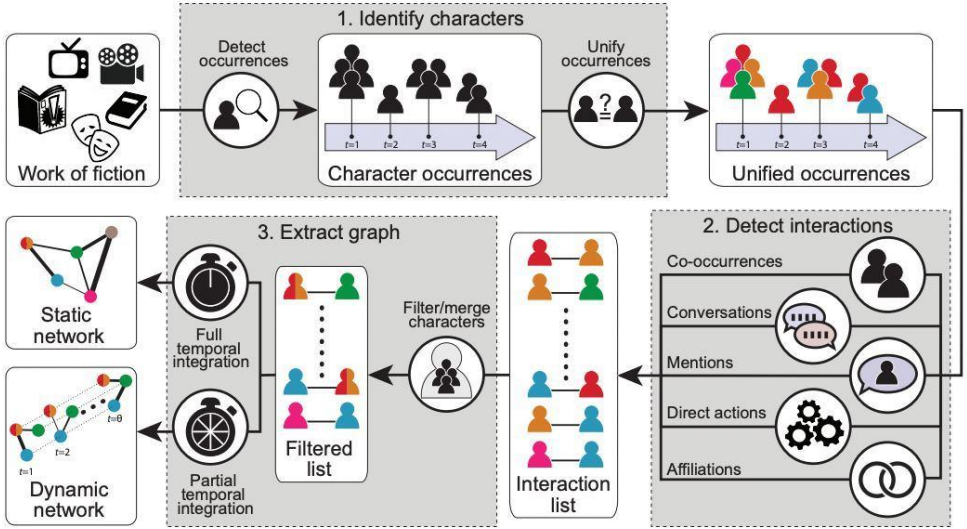
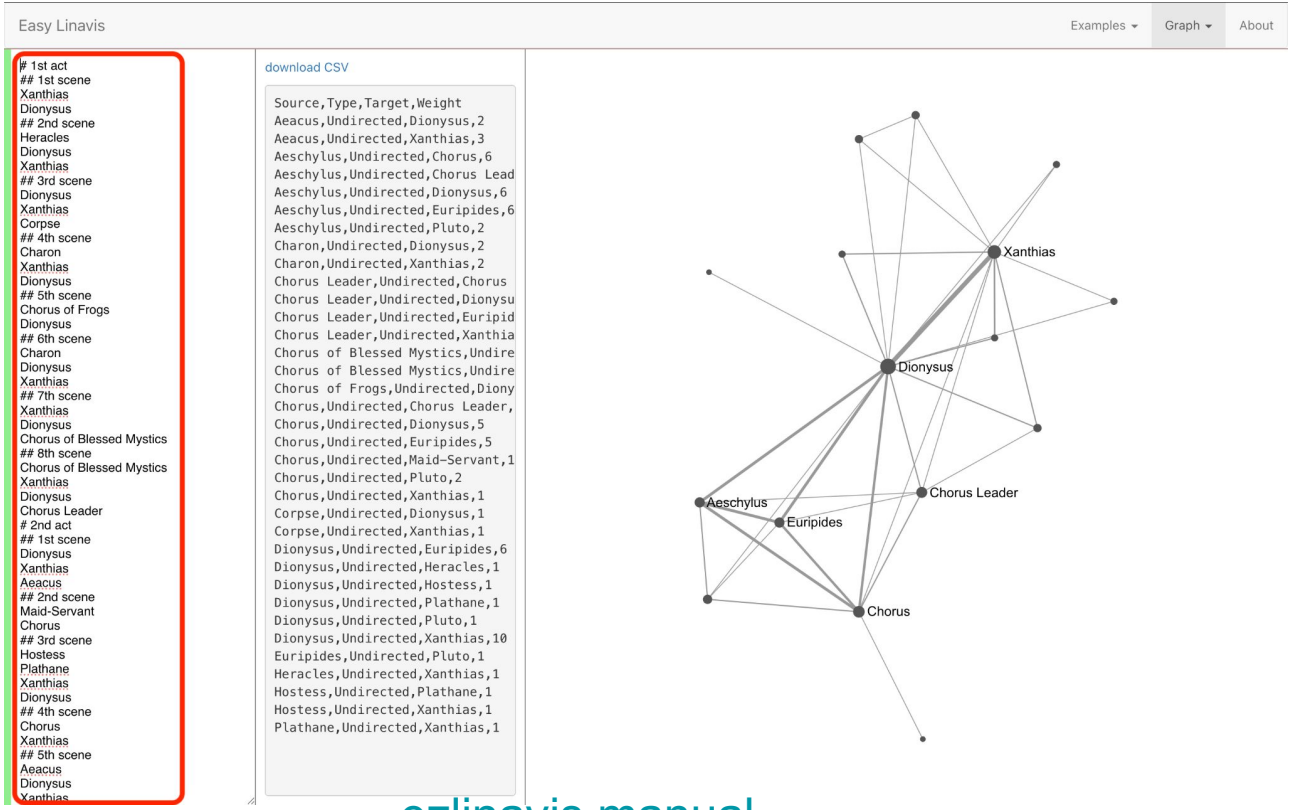


Fig. 1. Overview of the generic character network extraction process. Figure available at 10.6084/m9.figshare.7993040 under CC-BY license.

Labatut, Vincent, and Xavier Bost. 2019. 'Extraction and Analysis of Fictional Character Networks: A Survey'. ACM Computing Surveys 52 (5): 1–40. https://doi.org/10.1145/3344548.

Many encode by hand in a spreadsheet or in EzLinaVis:



ezlinavis manual

Let us play around: ezlinavis.dracor.org

(you can also find it on dracor.org under ‘Tools’ menu)

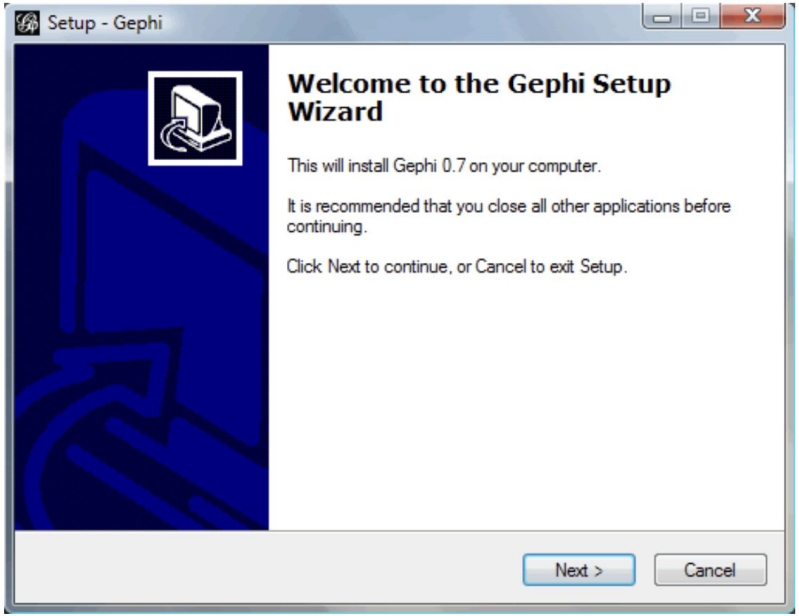
ezlinavis manual

Break! During the break you *may* start installing Gephi

gephi.org/users/install

Windows

1. After the download completes, run the installer and follow the steps.



Mac OS X

1. After the download completes, click on the downloaded *.dmg* file.
2. Drag the gephi application in your Application folder.

Linux

1. After the download completes, unzip and untar the file in a directory.
2. Run it by executing *./bin/gephi* script file.

Part 3: analyzing the data

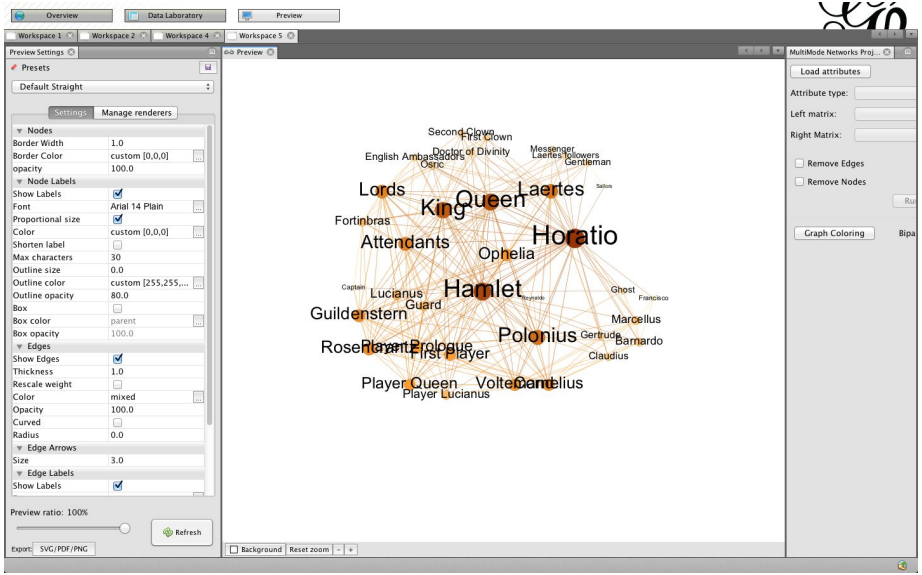
- Tools for network analysis
- Gephi
- Hands-on

Tools people use to analyse network:

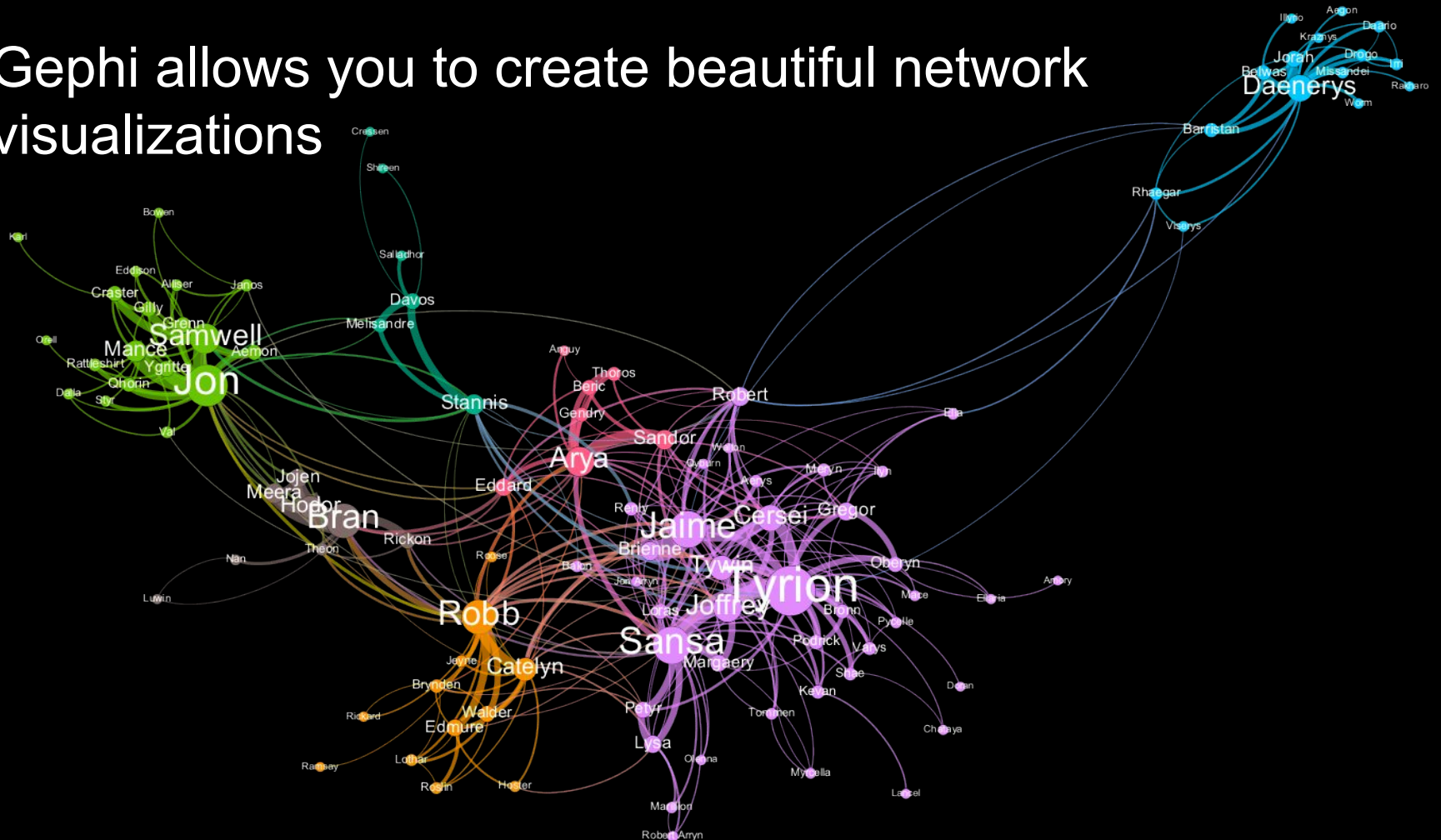
Gephi (gephi.org)

A tool for network visualization and analysis

"like Photoshop but for graph data" (gephi.org)



Gephi allows you to create beautiful network visualizations

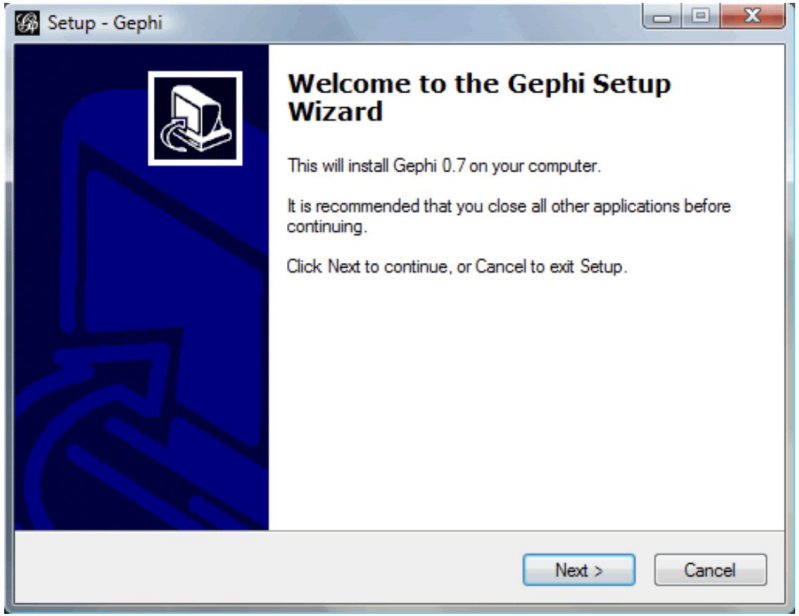


Installing Gephi

gephi.org/users/install

Windows

1. After the download completes, run the installer and follow the steps.



Mac OS X

1. After the download completes, click on the downloaded *.dmg* file.
2. Drag the gephi application in your Application folder.

Linux

1. After the download completes, unzip and untar the file in a directory.
2. Run it by executing *./bin/gephi* script file.

Hands-on with Gephi: visualizing a network following the instruction

Bonus: visualising a directed network Gephi

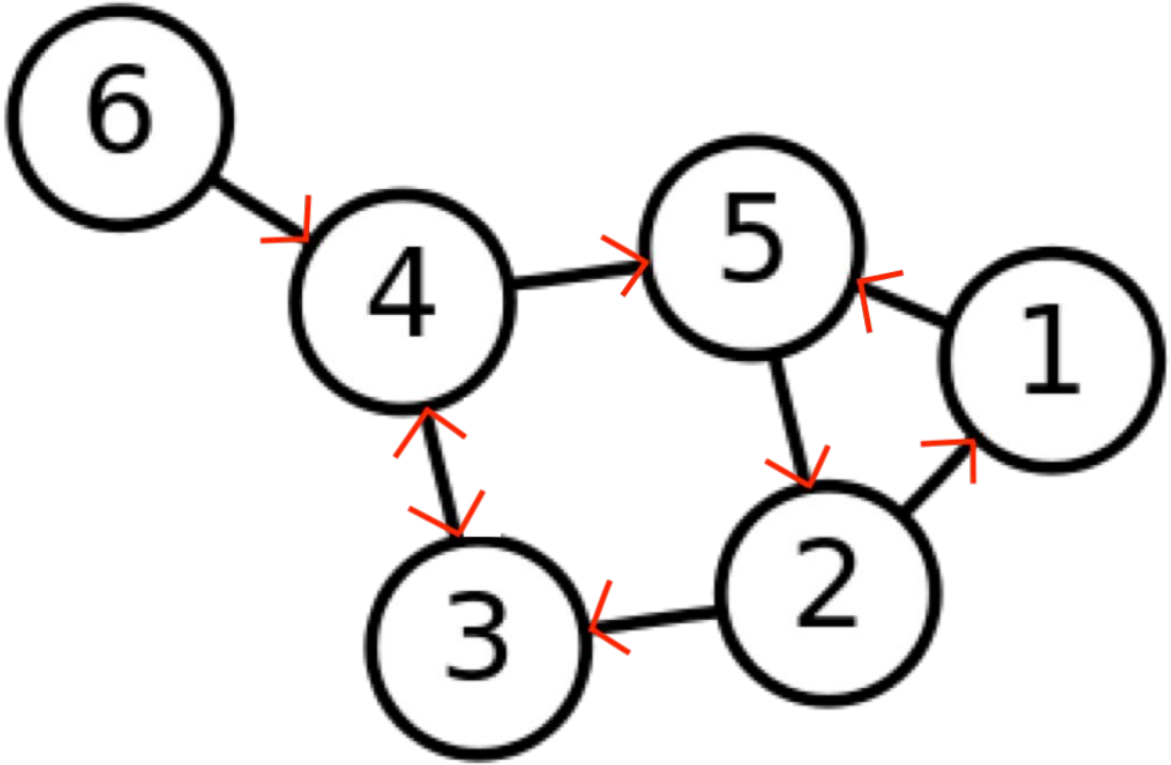
People of Medieval Scotland Project

poms.ac.uk/

Directed network data from the POMS project:

- Network (gexf-file 2 MB) of the so-called grantors and beneficiaries in medieval scotland
- Based on the information about «transactions» (donations, values transfer) of 2225 people or organisational entities (e.g. churches, abbeys) from 4063 Scottish medieval docs:
  - «*...explicitly made statements in the medieval sources about connections between individuals*».
  - *«...case study is based around the factoid type 'transaction'. The study incorporated only*
  - *transactions from the following document types: charter, charter/brieve, notification, agreement and settlement, because these for the most part contain evidence about dispositive transactions»*
  - *«There are 5351 transactions drawn from 4063 documents involving 2225 persons and institutions allowing for an average of 2.4 transactions per person»*
- Transactions are asymmetrical (i give to you IS NOT EQUAL you give to me), so the network is directed

Directed graph (network)



Directed network data from the POMS project:

- Network (gexf-file 2 MB) of the so-called grantors and beneficiaries in medieval scotland
- Based on the information about «transactions» (donations, values transfer) of 2225 people or organisational entities (e.g. churches, abbeys) from 4063 Scottish medieval docs:
  - «*...explicitly made statements in the medieval sources about connections between individuals*».
  - *«...case study is based around the factoid type 'transaction'. The study incorporated only*
  - *transactions from the following document types: charter, charter/brieve, notification, agreement and settlement, because these for the most part contain evidence about dispositive transactions»*
  - *«There are 5351 transactions drawn from 4063 documents involving 2225 persons and institutions allowing for an average of 2.4 transactions per person»*
- Transactions are asymmetrical (i give to you IS NOT EQUAL you give to me), so the network is directed

Directed network data from the POMS project:

- Transactions have some monetary value 💰💰💰 so the network is weighted
- More detailed decription of these data (chapter in a book about the ‘People of Medieval Scotland’ project):

3 Networks of Grantors & Beneficiaries

The three categories of relationship networks examined in chapter two reflected explicitly made statements in the medieval sources about connections between individuals. The following chapters examine social networks produced from other sorts of connections between medieval persons. Chapter three examines the links between grantors and beneficiaries of charters. Most of the documents in the database record gifts, confirmations, sales, and so forth, of land or other property from one person or institution to another. The enactment of these transactions set in train long-standing relationships. The anthropological and sociological literature on gift-giving and the social relationships engendered in gift-giving is vast. It is not the job of social network analysis to speculate on the exact nature of these relationships, rather to allow us access to these networks in ways which were hitherto impossible. While chapter two’s analyses were based on the factoid type ‘relationship’ in the PoMS database, chapter three’s case study is based around the factoid type ‘transaction’. The study incorporated only transactions from the following document types: charter, charter/brieve, notification, agreement and settlement, because these for the most part contain evidence about dispositive transactions, like gift-giving. The parameters of the study are as follows:

Table 3.1. Grantor and beneficiary study parameters

Number of documents	4063
Number of transactions	5351
Number of people/ institutions	2225
Average transactions per person	2.4

Thank you for your attention,
and apologies for talking too much :))