
Character network analysis of Émile Zola's *Les Rougon-Macquart*

DH 2015, Sydney
Yannick Rochat, DHLAB, EPFL

Character networks

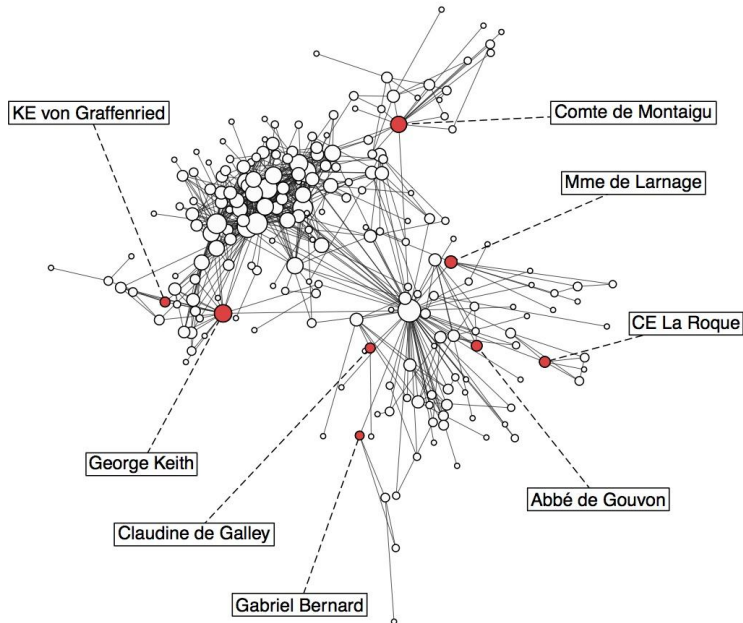
- Nodes symbolize characters
- Edges symbolize proximity in the discourse

=> ~~social~~ network of characters

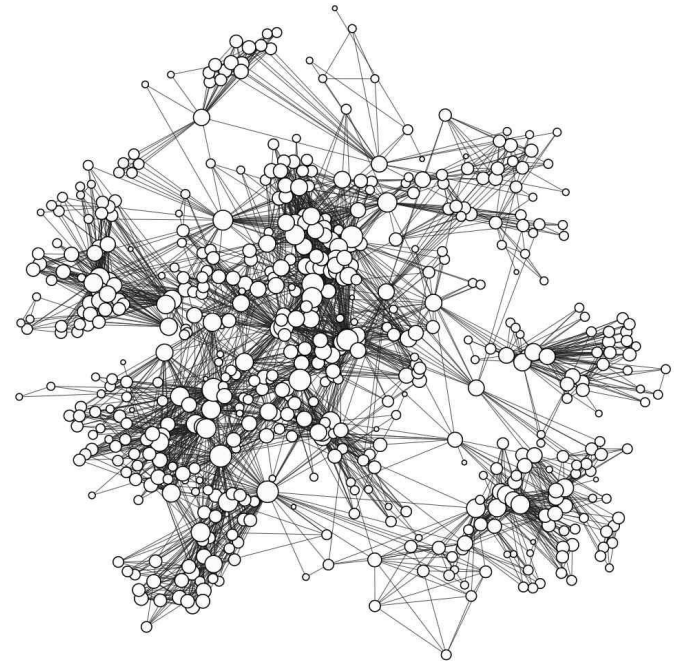
Character networks

- Most novels can be turned into networks
 - Different orders, different sizes
 - The network provides a signature of the novel
 - Application of measures
-

Character networks



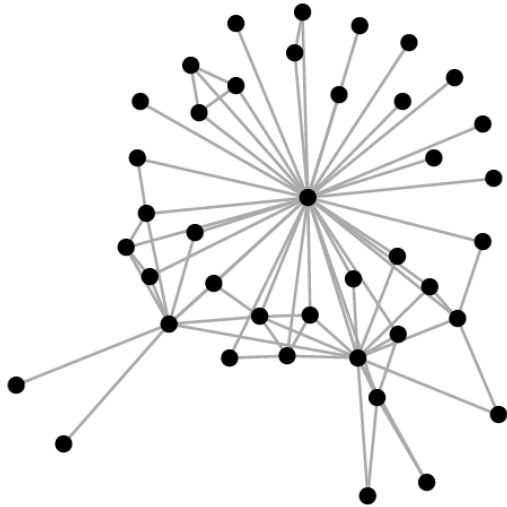
Les Confessions, J.-J. Rousseau



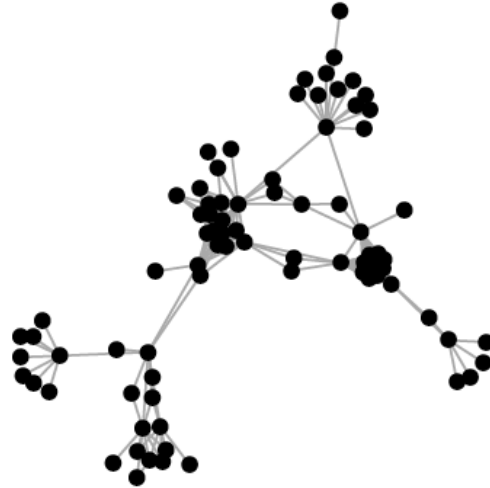
Sagas of Icelanders

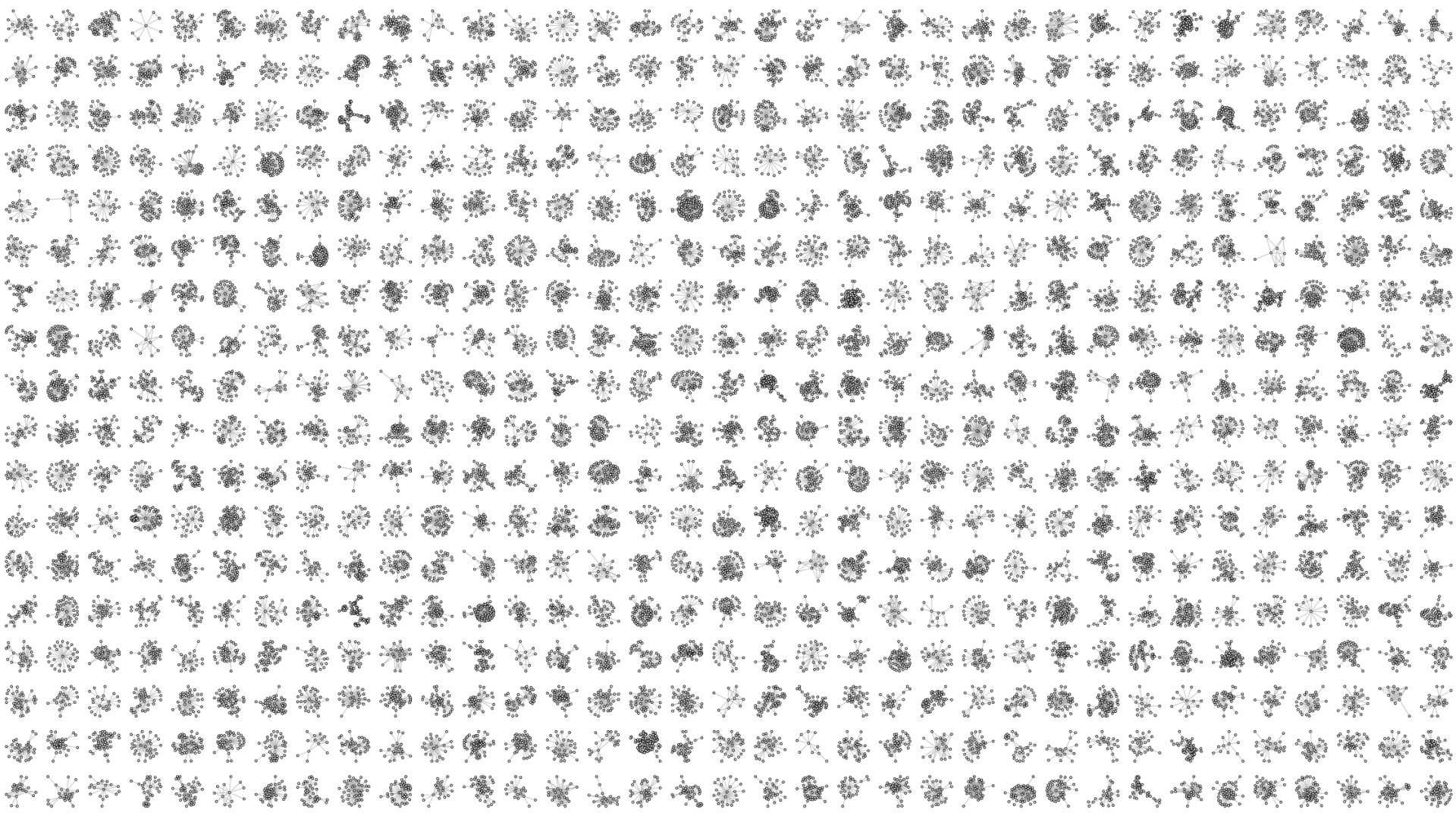
Character networks

Erin Brockovich



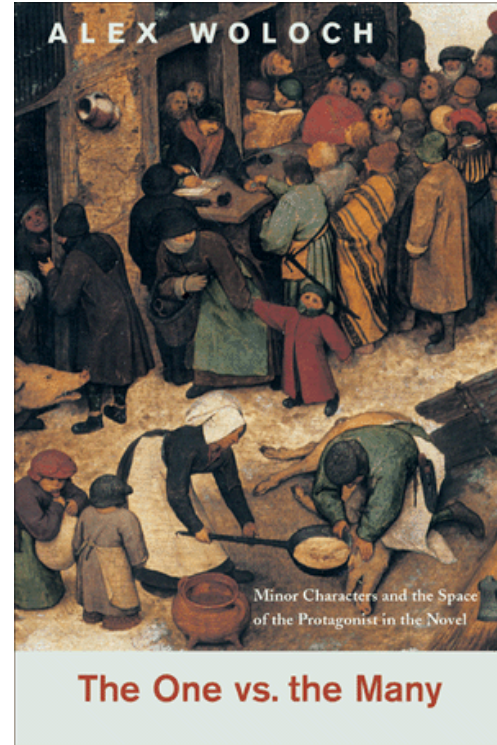
Magnolia



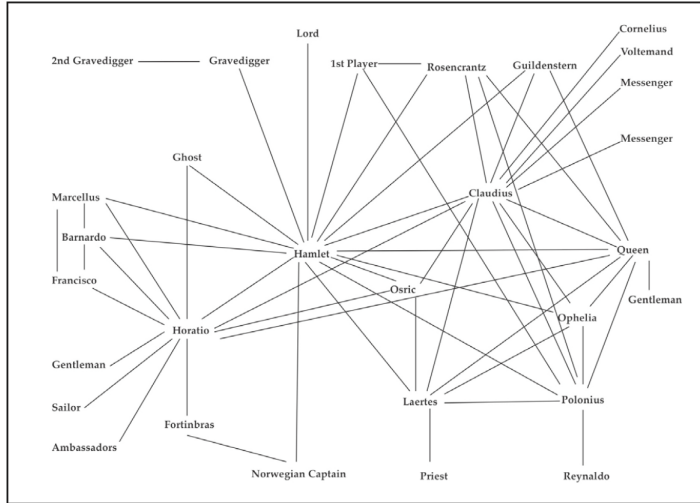


Character network analysis

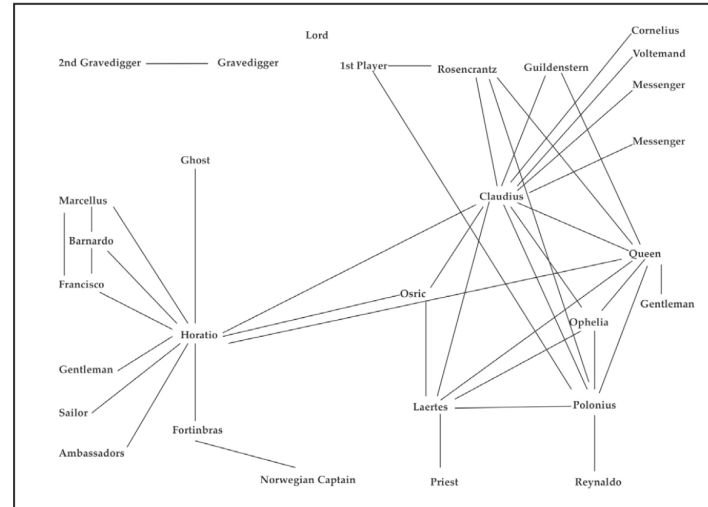
Character-space
Character-system
(Woloch, 2003)



Character network analysis



Plot analysis (Moretti, 2011)



Character network analysis

Statistical physics approach (Mac Carron & Kenna, 2012)

Network		N	$\langle k \rangle$	ℓ	ℓ_{rand}	ℓ_{max}	C	C_{rand}	G_c	r
Beowulf	All	74	4.45	2.37	2.88	6	0.69	0.06	50 (67.5%)	-0.10
	Hostile	31	1.67	2.08	3.25	4	0	0.05	10 (32.2%)	-0.20
	Friendly	68	4.12	2.45	2.98	6	0.69	0.06	45 (66.1%)	-0.03
Táin	All	404	6.10	2.76	3.32	7	0.82	0.02	398 (98.5%)	-0.33
	Hostile	144	2.33	2.93	5.88	7	0.17	0.02	131 (90.9%)	-0.36
	Friendly	385	5.67	2.84	3.43	7	0.84	0.01	350 (90.9%)	-0.32
Iliad	All	716	7.40	3.54	3.28	11	0.57	0.01	707 (98.7%)	-0.08
	Hostile	321	2.25	4.10	7.12	9	0	0.01	288 (89.4%)	-0.39
	Friendly	664	6.98	3.83	3.34	12	0.62	0.01	547 (82.3%)	0.10
Beowulf*	Friendly	67	3.49	2.83	3.36	7	0.68	0.05	43 (64.2%)	0.01
Táin*	Friendly	324	3.71	3.88	4.41	8	0.69	0.01	201 (62.0%)	0.04

	Social	Myth (friendly)	Fiction
Small world	Yes	Yes	Yes
Hierarchy	Yes	Yes	Yes
$p(k)$	Power law	Power law	Exp.
Scale free	Yes	Yes	No
G_c	< 90%	< 90%	> 90%
TA	Vulnerable	Vulnerable	Robust
RA	Robust	Robust	Robust
Assortative	Yes	Yes	No

Objectives

Extract features of novels.

Classify them based on :

- co-occurrences of characters
 - protagonists / minor characters
-

Les Rougon-Macquart

1802

INDEX GÉNÉRAL

L'Assommoir, II : 509, 510, 557, 615, 689 à 694, 708, 756 à 760, 779.

BIJARD (Henriette). Seconde fille des Bijard. — *L'Assommoir*, II : 557, 615, 691, 692, 694, 708, 757 à 760.

BIJARD (Jules). Troisième enfant des Bijard. — *L'Assommoir*, II : 615, 691, 692, 694, 708, 757 à 760.

BILLAUT (Auguste). Homme d'État français. — *Pot-Bouille*, III : 374.

BILLECOQ (Hermine). Protégée de Mme Correur. — *Son Excellence Eugène Rougon*, II : 56, 198, 200, 228, 249 à 252, 264, 270 à 272, 307.

BISMARCK (Otto de). Homme d'État prussien. — *Nana*, II : 1147, 1148 à 1152, 1163, 1164, 1178, 1179, 1483. — *L'Argent*, V : 180, 181, 191, 254, 259, 295, 296, 331. — *La Débâcle*, V : 437, 438, 440, 466, 586, 587, 601, 713, 725, 726, 808, 820, 858, 866.

BLANCHET. Député. — *Son Excellence Eugène Rougon*, II : 15.

BLAISOT. Banquier à Paris. — *L'Argent*, V : 128.

BLÉRIOT (de). Préfet. Réprime l'insurrection de la région de Plissans en décembre 1851. — *La Fortune des Rougon*, I : 212, 219, 291 à 293.

BLEUZE. Corderie à Montsou. — *Germinal*, III : 1137, 1461.

BLOND (Maria). Actrice du Théâtre des Folies-Dramatiques. — *Nana*, II : 1159, 1168 à 1193, 1203, 1246, 1247, 1250 à 1256, 1258, 1302, 1361, 1381, 1388, 1390,

1392, 1400, 1405, 1451, 1479 à 1484.

BOCHE. Concierge de l'immeuble habité par les Coupeau, rue de la Goutte-d'Or. — *L'Assommoir*, II : 387, 394, 433 à 461, 492 à 496, 499 à 502, 507, 520 à 522, 524, 552, 556, 559, 560, 566 à 569, 571 à 595, 599, 603 à 607, 609, 611, 630, 643, 646, 657, 660, 662, 667, 670, 672, 674, 675, 678, 679, 682 à 684, 696, 702, 708, 713, 725, 728, 731, 744, 750, 751, 754, 756, 771, 784, 786, 789, 790, 794, 795.

BOCHE (Mme). Épouse du précédent. — *L'Assommoir*, II : 378, 379, 382, 387 à 395, 398 à 401, 433, 434, 451, 454 à 460, 472, 479 à 483, 493 à 495, 499, 500, 502, 505, 507, 510, 513, 520 à 522, 524, 552, 556, 559 à 562, 566, 569, 571 à 595, 599, 601, 603, 611, 612, 630, 633, 643, 646, 657, 662, 667, 674, 675, 678, 679, 681, 682, 684, 696, 701, 708, 725, 728, 731, 744, 750, 751, 778, 784, 789, 790, 794. — *Nana*, II : 1366.

BOCHE (Pauline). Fille des précédents. — *L'Assommoir*, II : 494, 495, 519, 520, 559, 573, 575, 582, 592 à 594, 678 à 681, 711 à 715, 744.

BOCQUET (Clarisse). Maîtresse de Duveyrier. — *Pot-Bouille*, III : 95, 117, 124, 131 à 136, 150, 183, 192 à 196, 199, 206, 208, 210, 262, 275, 297, 302, 304, 306, 307, 309 à 312, 343, 350, 351, 355, 356.

BOCQUET (famille). Parents de la précédente. — *Pot-Bouille*, III : 131, 306, 307, 308 à 311, 350, 355.

DES PERSONNAGES

1803

BODIN (docteur). Vieux médecin de Passy. — *Une Page d'Amour*, II : 803, 809, 811, 822, 832, 845, 932, 950, 1048, 1049, 1056, 1060 à 1063, 1069 à 1071.

BOHAIN (marquis de). Membre du conseil d'administration de la Banque Universelle. — *L'Argent*, V : 100 à 104, 107, 121, 131 à 134, 165, 168, 199, 248, 249, 279, 300, 306, 311, 325, 331, 351, 355, 382.

*BOISSIER. Confiseur. — *Nana*, II : 1127, 1160.

*BONAPARTE (Napoléon). Empereur sous le nom de Napoléon I^{er}. — *La Fortune des Rougon*, I : 78, 82, 113, 315. — *La Curée*, I : 391, 395. — *Nana*, II : 1149. — *Germinal*, III : 1198. — *La Terre*, IV : 433. — *L'Argent*, V : 78, 231, 253, 383, 387. — *La Débâcle*, V : 415, 447 à 449, 453. — *Le Docteur Pascal*, V : 1010.

*BONAPARTE (Louis-Napoléon). Empereur sous le nom de Napoléon III. — *La Fortune des Rougon*, I : 75, 82, 89, 90, 93, 94, 101, 102, 105, 106, 254, 259, 265, 267, 304, 314. — *La Curée*, I : 342, 345, 395, 439, 440, 564, 596, 597. — *Le Ventre de Paris*, I : 661, 662, 704, 710, 758, 842, 850. — *Son Excellence Eugène Rougon*, II : 17, 18, 21, 27, 28, 32, 42 à 44, 49, 78, 80, 84, 88, 89, 97, 98, 102 à 106, 111, 133, 150 à 141, 150, 153 à 175, 179, 182 à 185, 202, 207, 208, 210, 211, 214, 217, 218, 221, 241, 242, 251, 252, 276, 278 à 296, 298, 318, 327, 332, 334, 345, 347, 353, 354, 367. — *L'Assom-*

moir, II : 580, 604 à 606, 626, 650, 734, 735, 737. — *Nana*, II : 1114, 1164, 1369, 1483. — *Pot-Bouille*, III : 87, 157, 220, 333, 374. — *An Bonheur des Dames*, III : 569. — *Germinal*, III : 1136, 1149, 1356, 1357, 1358. — *L'Œuvre*, IV : 118, 135, 273. — *La Terre*, IV : 415, 416, 433, 488, 499, 501, 681. — *La Bête Humaine*, IV : 1076, 1315, 1316. — *L'Argent*, V : 15, 16, 92, 94, 110, 179 à 181, 192, 197, 199, 210, 221, 231, 254 à 256, 277. — *La Débâcle*, V : 409, 413, 438 à 440, 443 à 447, 453, 454, 457 à 459, 464, 468, 471, 490, 493 à 499, 503, 520, 523, 547, 550, 551, 562, 565, 579, 580, 608, 609, 612, 616, 618, 621, 667 à 669, 682 à 684, 687, 688, 706, 716, 717, 724 à 727, 753, 754, 797, 801, 851, 859 à 861. — *Le Docteur Pascal*, V : 928.

*BONAPARTE (Napoléon-Louis). Fils de Napoléon III et de l'impératrice Eugène. Dit le Prince Impérial. — *Son Excellence Eugène Rougon*, II : 89, 96 à 98, 100, 102, 105, 158, 169, 221, 352. — *Germinal*, III : 1260. — *La Débâcle*, V : 410, 498, 580.

BONGRAND. Grand peintre. — *L'Œuvre*, IV : 86 à 89, 118, 119, 125, 127, 133, 134, 137, 180 à 188, 235, 274 à 276, 279, 281, 288 à 292, 301, 353 à 363.

BONHOMME (Jacques). Type mythique du paysan. — *La Terre*, IV : 428 à 436.

BONNAUD. Ancien chef de la comptabilité aux Chemins de fer du Nord. — *Pot-Bouille*, III : 49, 54.

Les Rougon-Macquart

1. OCR (5min)
 2. Manual/regex cleaning (10-15h)
 3. Importation into R/igraph
 4. Remove most obvious artefacts
-

First step: pages

Enumeration of co-occurrences on couples of adjacent pages.

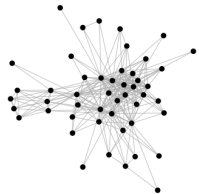
Same page gives 2 such co-occurrences : $[n-1, n]$ & $[n, n+1]$.

Second step: threshold

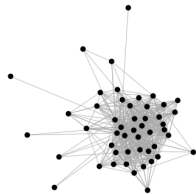
(#co-occurrences $>$ threshold T) \Rightarrow relation.

In practice, we consider : $T = 3$.

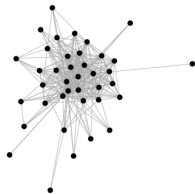
La Fortune des Rougon



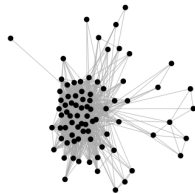
La Curée



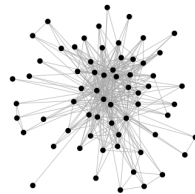
Le Ventre de Paris



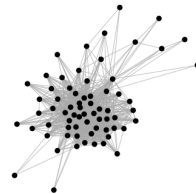
Germinal



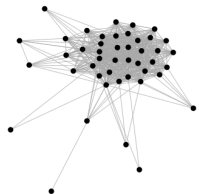
L'Oeuvre



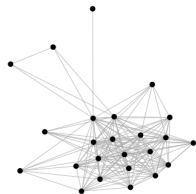
La Terre



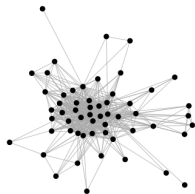
La Conquête de Plassans



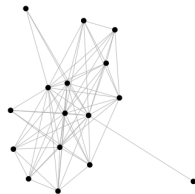
La Faute de l'Abbé Mouret



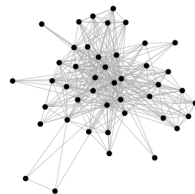
Son Excellence Eugène Rougon



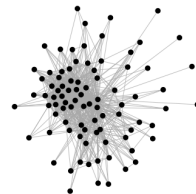
Le Rêve



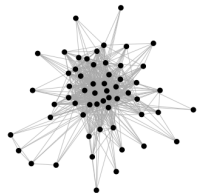
La Bête Humaine



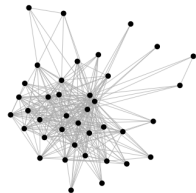
L'Argent



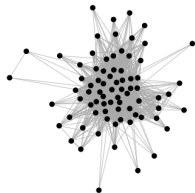
L'Assommoir



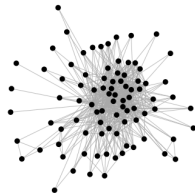
Une Page d'Amour



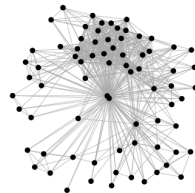
Nana



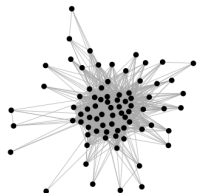
La Débâcle



Le Docteur Pascal



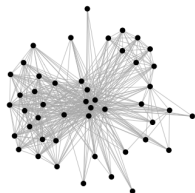
Pot-Bouille



Au Bonheur des Dames



La Joie de Vivre



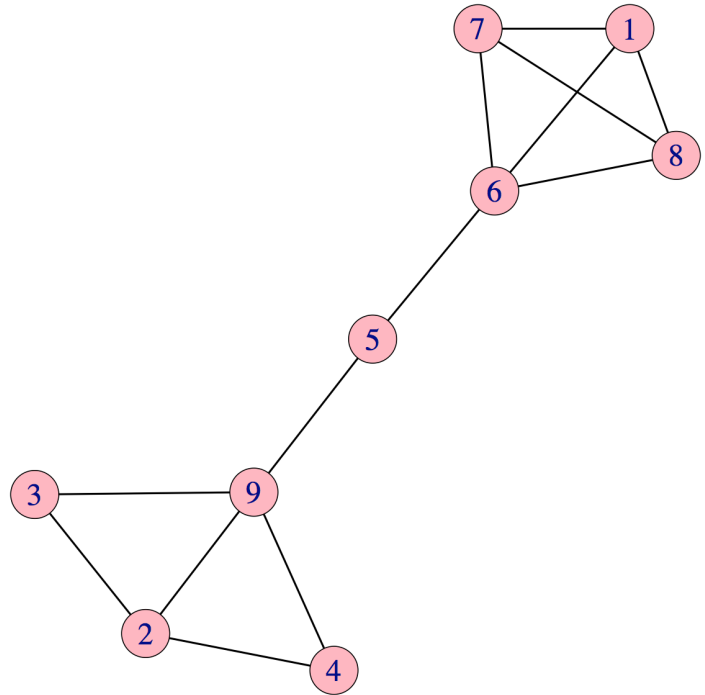
Betweenness centralization

To determine the overall centrality of a point, p_k , we need merely to sum its partial betweenness values for all unordered pairs of points where $i \neq j \neq k$:

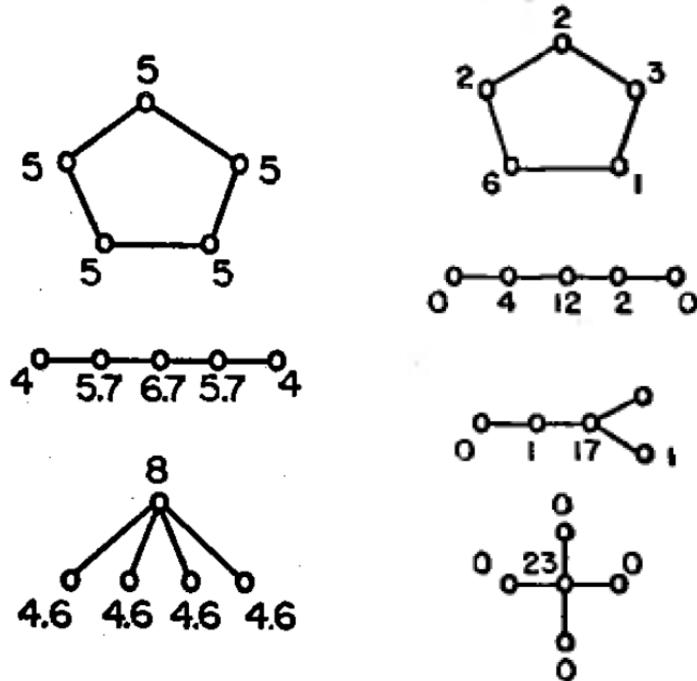
$$C_B(p_k) = \sum_{i < j} b_{ij}(p_k),$$

where n = the number of points in the graph.

Freeman, 1977



Betweenness centralization



Then a natural measure of the dominance of the most central point is

$$C'_B = \frac{\sum_{i=1}^n [C_B(p_k^*) - C'_B(p_i)]}{n-1},$$

which is the average difference in centrality between the most central point and all others.

Freeman, 1977

Betweenness centralization

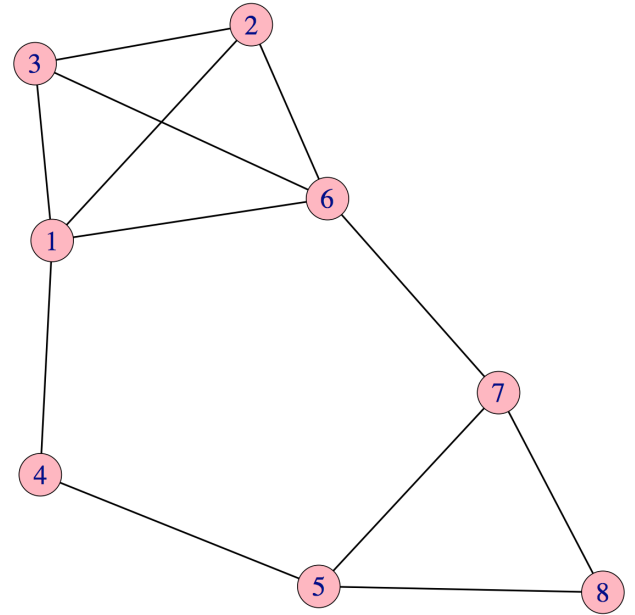
Novel	C_B
La Fortune des Rougon	0.142
La Curée	0.105
Le Ventre de Paris	0.167
La Conquête de Plassans	0.13
La Faute de l'Abbé Mouret	0.187
Son Excellence Eugène Rougon	0.271
L'Assommoir	0.185
Une Page d'Amour	0.162
Nana	0.127
Pot-Bouille	0.126

Novel	C_B
Au Bonheur des Dames	0.159
La Joie de Vivre	0.088
Germinal	0.245
L'Oeuvre	0.412
La Terre	0.097
Le Rêve	0.208
La Bête Humaine	0.141
L'Argent	0.364
La Débâcle	0.317
Le Docteur Pascal	0.283

Coreness

“The **k-core** of a graph is a maximal subgraph in which each vertex has at least degree k .

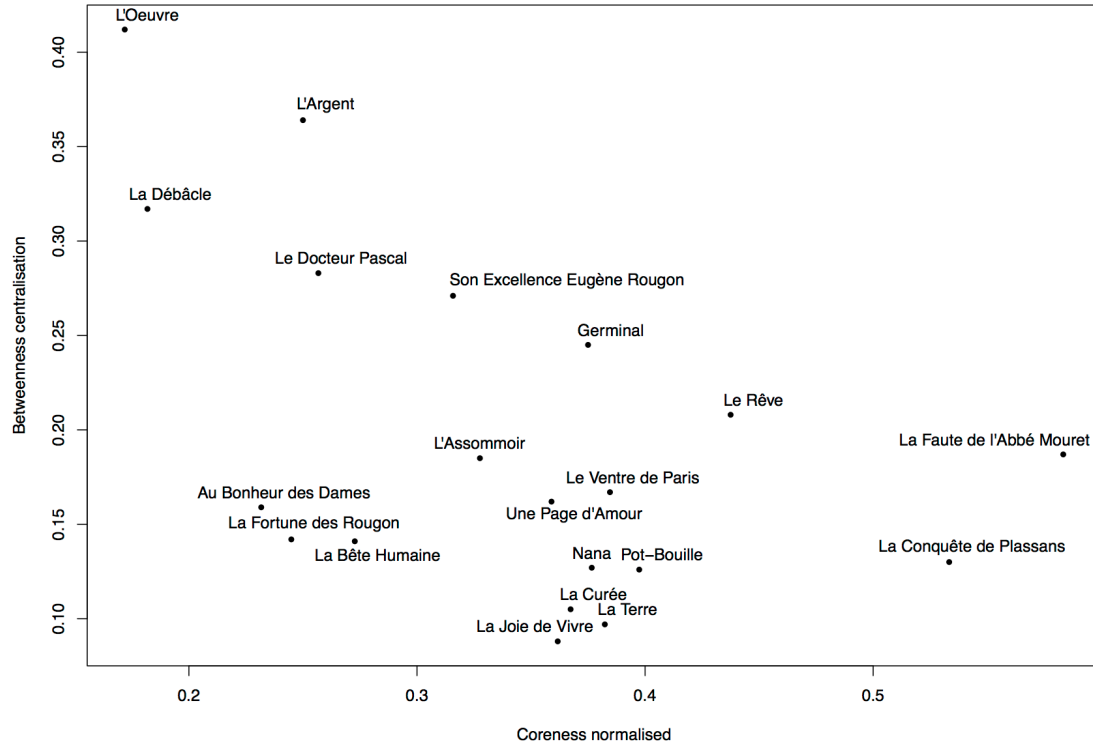
The **coreness** of a vertex is k if it belongs to the k -core but not to the $(k+1)$ -core.” (Csardi, 2006)



Coreness

Novel	Order	Coreness max	Coreness norm.
La Fortune des Rougon	49	12	0.24
La Curée	49	18	0.37
Le Ventre de Paris	39	15	0.39
La Conquête de Plassans	45	24	0.53
La Faute de l'Abbé Mouret	24	14	0.58
Son Excellence Eugène Rougon	57	18	0.32
L'Assommoir	58	19	0.33
Une Page d'Amour	39	14	0.36
Nana	77	29	0.38
Pot-Bouille	78	31	0.40
Au Bonheur des Dames	82	19	0.23
La Joie de Vivre	47	17	0.36
Germinal	80	30	0.38
L'Oeuvre	64	11	0.17
La Terre	68	26	0.38
Le Rêve	16	7	0.44
La Bête Humaine	44	12	0.27
L'Argent	88	22	0.25
La Débâcle	88	16	0.18
Le Docteur Pascal	74	19	0.26

Coreness



Conclusion

Classification through main protagonist and cores of protagonists.

On this corpus, coreness and betweenness centralisation highlight different properties.

Diversity within Zola's narration.

Future works

Take into account weights of edges for a better accuracy of “coreness”.

How do we study interlocking novels ? (Zola, Balzac, ...)

Representation and study of temporality.

Thank you!

