A Network of Mathematical Theorems

A.Chotai, C.Mavroyiakoumou, I.Papadopoulos, X.Wen

Imperial College London

Department of Mathematics

June 17, 2015

A Network of Theorems

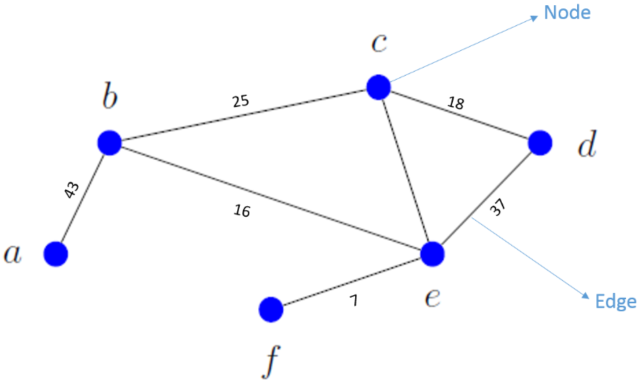


Fig: A Weighted, Undirected Network

A Network of Theorems

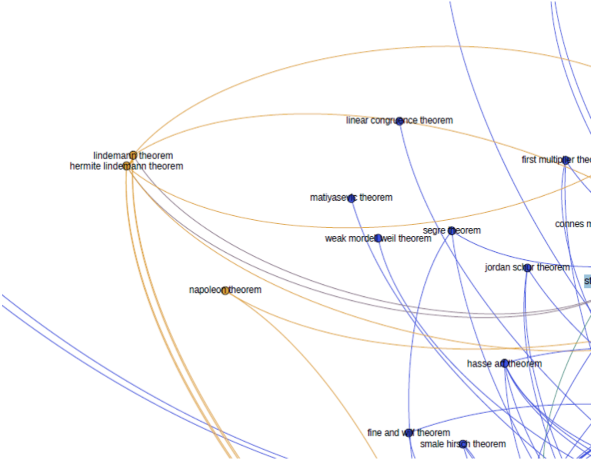
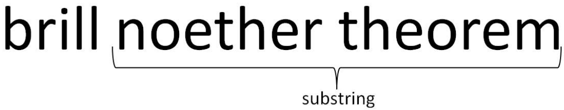


Fig: A Part of Our Network of Theorems

The Substring Bug



- Removal of 78 nodes and approximately 5000 edges
- We keep any nodes that the substring bug contributes less than 5% to the frequency e.g. the central limit theorem which is a substring of Lindeberg central limit theorem.

Removal of duplicates

Duplicate nodes correspond to theorems that have the same name and appear in our data set with

different

- ID number

Removal of duplicates

Duplicate nodes correspond to theorems that have the same name and appear in our data set with

different

- ID number

but same

- Frequency
- Degree

Theorem IDs	Theorem Name	Frequency	Degree
118, 167	distortion theorem	295	80
143, 868	Helmholtz theorem	190	40
175, 767	Torelli theorem	82	36
191, 362	Godel incompleteness theorem	26	16
219, 698	Bertini theorem	78	42
253, 547	art gallery theorem	179	12
272, 531	Fredholm theorem	134	89
426, 1460	Chasles theorem	60	15
553, 1619	Dirac theorem	27	17
571, 749	Euclid theorem	7	5
766, 1580	Foster theorem	7	4
836, 1644	Oka theorem	8	5

Removal of duplicates

Duplicate nodes correspond to theorems that have the same name and appear in our data set with

different

- ID number

but same

- Frequency
- Degree

Theorem IDs	Theorem Name	Frequency	Degree
118, 167	distortion theorem	295	80
143, 868	Helmholtz theorem	190	40
175, 767	Torelli theorem	82	36
191, 362	Godel incompleteness theorem	26	16
219, 698	Bertini theorem	78	42
253, 547	art gallery theorem	179	12
272, 531	Fredholm theorem	134	89
426, 1460	Chasles theorem	60	15
553, 1619	Dirac theorem	27	17
571, 749	Euclid theorem	7	5
766, 1580	Foster theorem	7	4
836, 1644	Oka theorem	8	5

Action taken: Delete the 12 duplicates from our data set!

Compare initial n consecutive characters

Potential bugs that arise from the $n = 7$ case include:

Compare initial n consecutive characters

Potential bugs that arise from the $n = 7$ case include:

- Theorems that have **various spellings**.
  - e.g. *Pythagoras theorem*, *Pythagoreas theorem* and *Pythagorean theorem*

Compare initial n consecutive characters

Potential bugs that arise from the $n = 7$ case include:

- Theorems that have **various spellings**.
  - e.g. *Pythagoras theorem*, *Pythagoreas theorem* and *Pythagorean theorem*
- Theorems that are the same but when mentioned in a publication often have **one of the contributors' name omitted**.
  - e.g. *Ascoli theorem* and *Ascoli Arzela theorem*

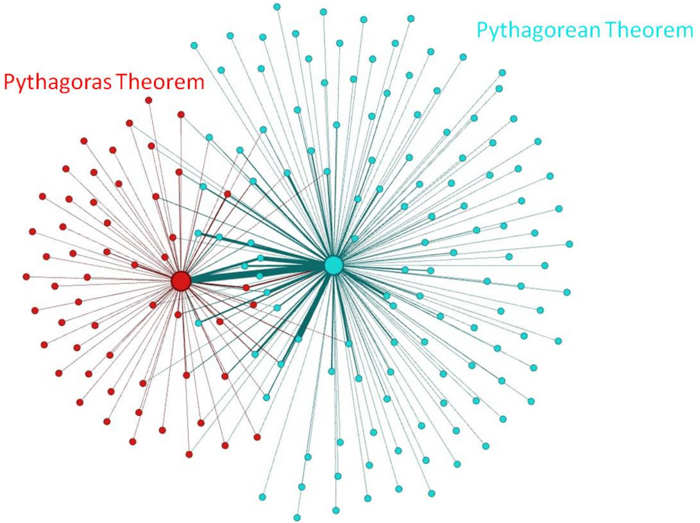
Compare initial n consecutive characters

Potential bugs that arise from the $n = 7$ case include:

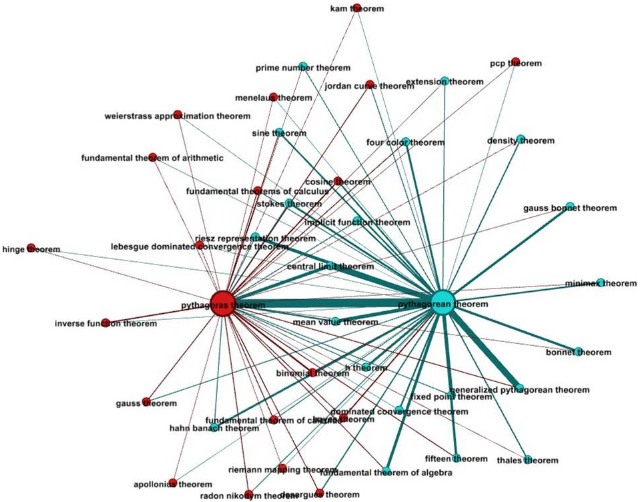
- Theorems that have **various spellings**.
  - e.g. *Pythagoras theorem*, *Pythagoreas theorem* and *Pythagorean theorem*
- Theorems that are the same but when mentioned in a publication often have **one of the contributors' name omitted**.
  - e.g. *Ascoli theorem* and *Ascoli Arzela theorem*
- Nodes that describe a **general class of theorems** rather than being an actual theorem.
  - e.g. *theorems on continuation*, *theorems on sums of squares*, and *theorems of Euclid*

This method motivates the bug hunting.

Merging Nodes



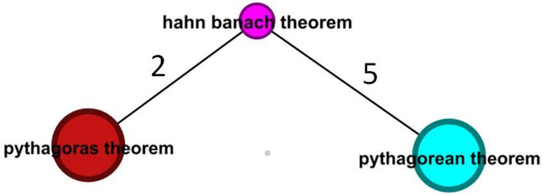
Merging Nodes

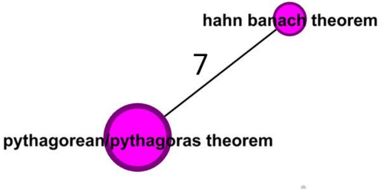


Merging Nodes

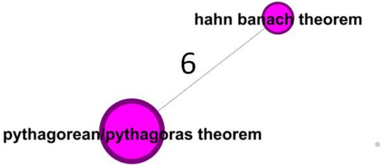
Sum the weights of the two edges to a shared node

- Pythagoras theorem - Hahn Banach theorem (2)
- Pythagorean theorem - Hahn Banach theorem (3)



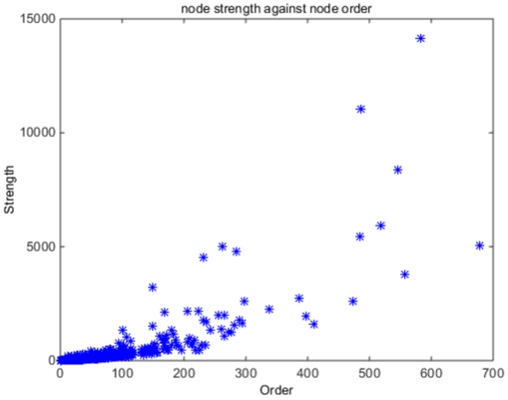


$\left(1 - \frac{11}{40}\right) \times 7 = 5.075$. We round this up to 6.

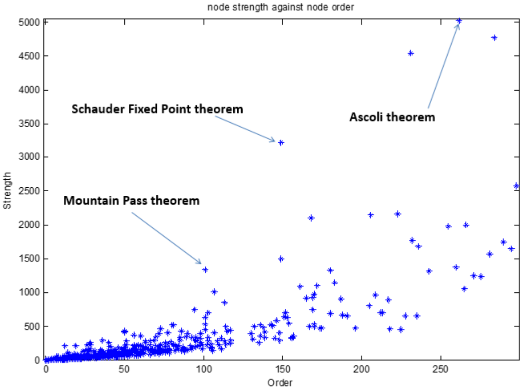


Order and Strength

- The order of a node is the number of different theorems to which the node is connected
- The strength of a node is the sum of the weights of each edge incident to that node

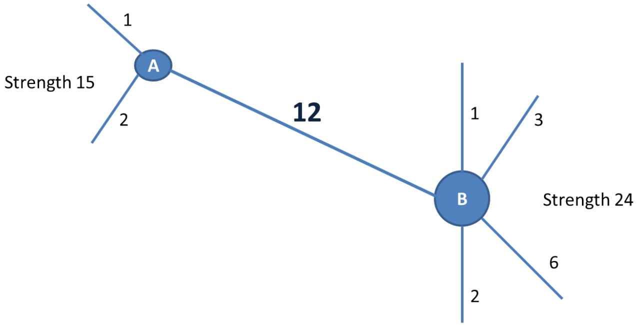


We are interested in nodes with low degrees and high strengths.



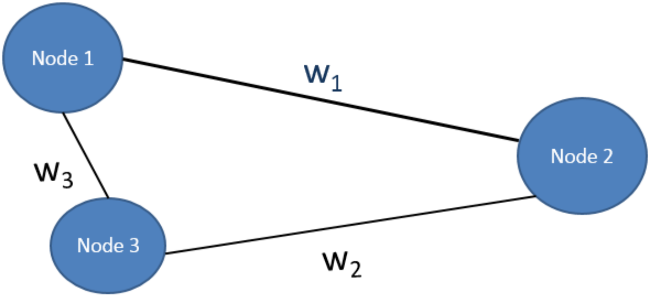
Normalising Weights

It is helpful to consider the nodes at a local level.



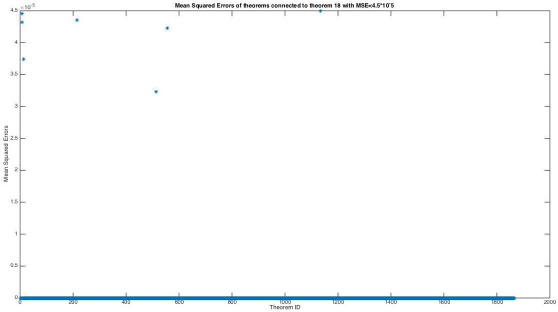
We create a non-symmetric matrix P of normalised weights, where $P_{ij} = \frac{w_{ij}}{s_i}$

Mean Squared Errors



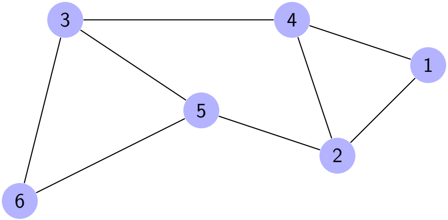
$$P = \begin{pmatrix} 0 & P_{12} & P_{13} \\ P_{21} & 0 & P_{23} \\ P_{31} & P_{32} & 0 \end{pmatrix}$$

Mean Squared Errors with Green's theorem



Theorem Name
h-theorem
divergence theorem
Stoke's theorem
Liouville theorem
Goursat theorem
Alexandrov theorem

Shortest Paths



In this graph, the shortest path from 1 to 3 is through 4.

The diameter is the longest length of the shortest paths.

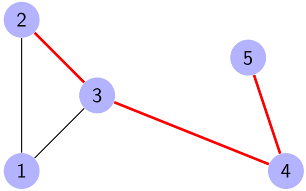


Fig 1: This network has diameter 3.

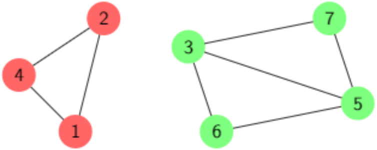


Fig 2: This network has infinite diameter

The four couples

Theorem 1 ID	Theorem 1 name	Theorem 2 ID	Theorem 2 name
936	Karp Lipton theorem	1250	Sipser Lautemann theorem
1144	Lickorish Twist theorem	1253	Lickorish Wallace theorem
1214	Mohr Mascheroni theorem	1267	Poncelet Steiner theorem
1264	Mertens First theorem	1288	Mertens Second theorem

Fig: The four distinct pairs of theorems that we stripped off

Node Betweenness Centrality

Which theorem is 'important'?

Node Betweenness Centrality

Which theorem is 'important'? Look at its node betweenness centrality

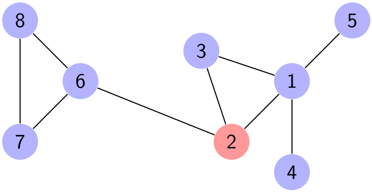
Node Betweenness Centrality

Which theorem is ‘important’? Look at its node betweenness centrality

Definition

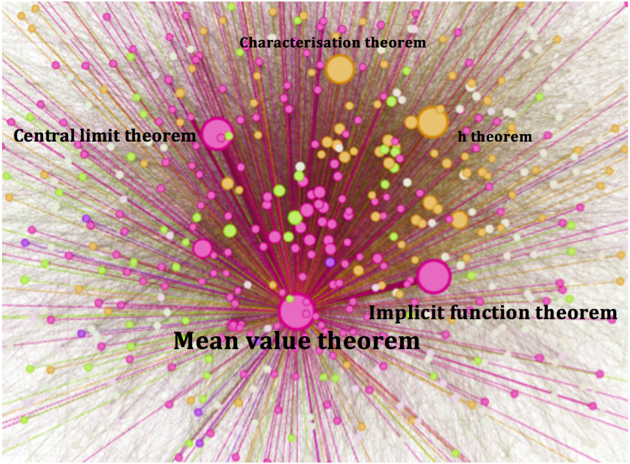
$$BC(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}}$$

where $\sigma_{st}(n)$ is the total number of shortest paths passing through node n and σ_{st} is the total number of shortest paths from node s to node t.



In this graph, node 2 has the highest node betweenness centrality.

MVT most 'important' intermediary



Theorems with the highest node betweenness centrality.[1]

[1] Diagram produced in *Gephi*

Normalised Node Betweenness Centrality

Normalised $BC(n)$

$$BC(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}} \div \binom{N-1}{2}$$

where N is the total number of nodes.

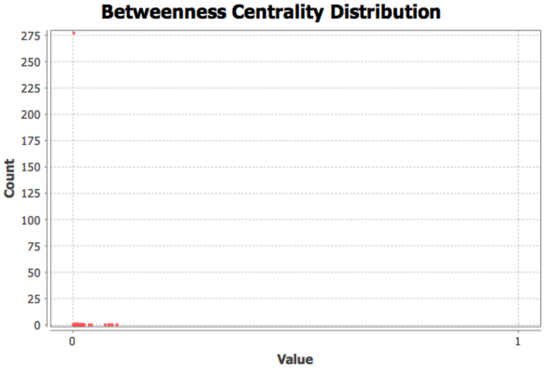
Normalised Node Betweenness Centrality

Normalised $BC(n)$

$$BC(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}} \div \binom{N-1}{2}$$

where N is the total number of nodes.

278 nodes have zero node betweenness centrality, which means that they lie on the boundary of the network of mathematical theorems.



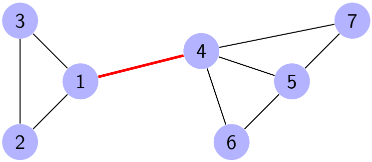
Edge Betweenness Centrality

Edge betweenness centrality indicates the 'importance' of a edge.

Definition

$$BC(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

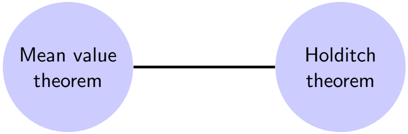
where $\sigma_{st}(e)$ is the total number of shortest paths going through edge e and σ_{st} is the total number of shortest paths from node s to node t.



The edge connecting 1 to 4 has the highest edge betweenness centrality.

Most 'important' pair of theorems

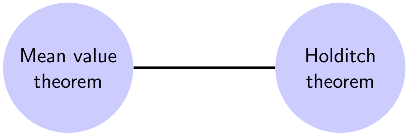
This pair of theorems has the highest edge betweenness centrality



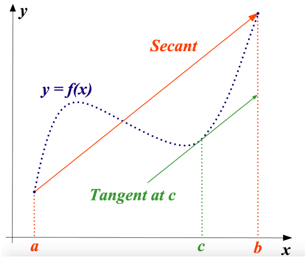
This edge lies on most shortest paths between other theorems.

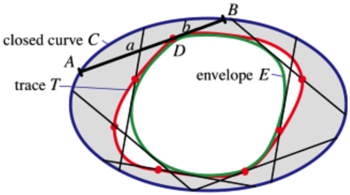
Most 'important' pair of theorems

This pair of theorems has the highest edge betweenness centrality



This edge lies on most shortest paths between other theorems.



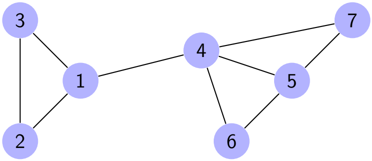


Girvan-Newman Algorithm

Reveals community structure of the network

Algorithm 1 Girvan – Newman Algorithm

1: Calculate the edge betweenness centrality of all the edges.

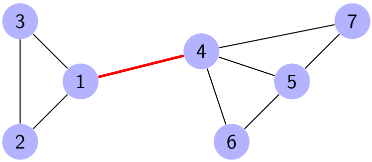


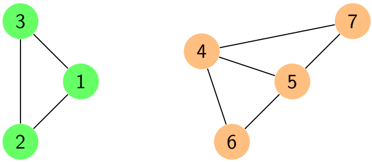
Girvan-Newman Algorithm

Reveals community structure of the network

Algorithm 1 Girvan – Newman Algorithm

1: Calculate the edge betweenness centrality of all the edges.
2: Remove the edge with the highest edge betweenness centrality.



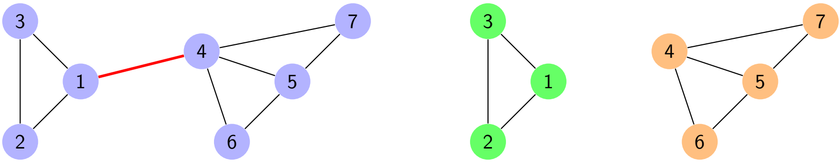


Girvan-Newman Algorithm

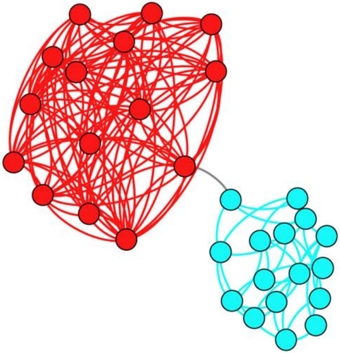
Reveals community structure of the network

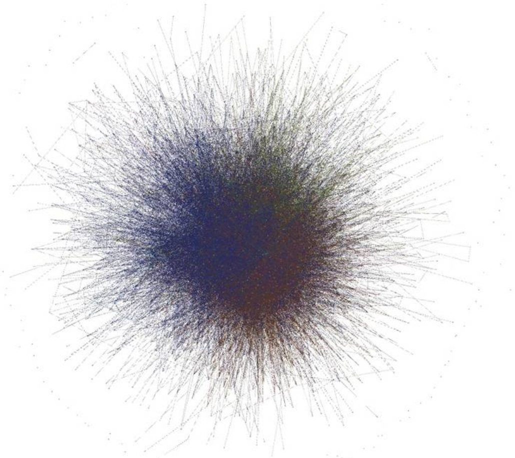
Algorithm 1 Girvan – Newman Algorithm

1: Calculate the edge betweenness centrality of all the edges.
2: Remove the edge with the highest edge betweenness centrality.
3: Recalculate the edge betweenness centrality for the remaining edges.
4: Repeat step 2 and step 3 until all the edges have been deleted.

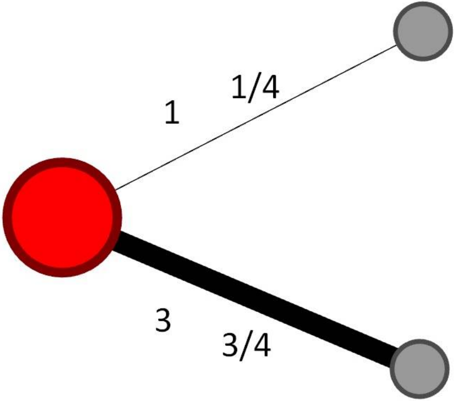


Community Detection





Stability



Community Results

Community (3.#)	Count	Percentage
0	492	37.70%
1	735	56.32%
2	55	4.21%
3	2	0.15%
4	13	1.00%
5	2	0.15%
6	2	0.15%
7	2	0.15%
8	2	0.15%

Normalised Laplacian after deleting central limit theorem, h theorem,divergence theorem, mean value theorem, bayes theorem, implicit function theorem, fluctuation dissipation theorem, gauss theorem and characterisation theorem

Theorem ID	Theorem name	Frequency
84	final value theorem	483
208	initial value theorem	165
240	gershgorin theorem	111
369	gershgorin circle theorem	90
424	gerschgorin circle theorem	111
506	gauss lucas theorem	40
544	bauer fike theorem	65
736	poincare separation theorem	29
944	gershgorin disc theorem	30
1108	bisectors theorem	6
1132	gerschgorin disc theorem	12
11169	perpendicular bisector theorem	2
1266	sylvester determinant theorem	2

Community Results

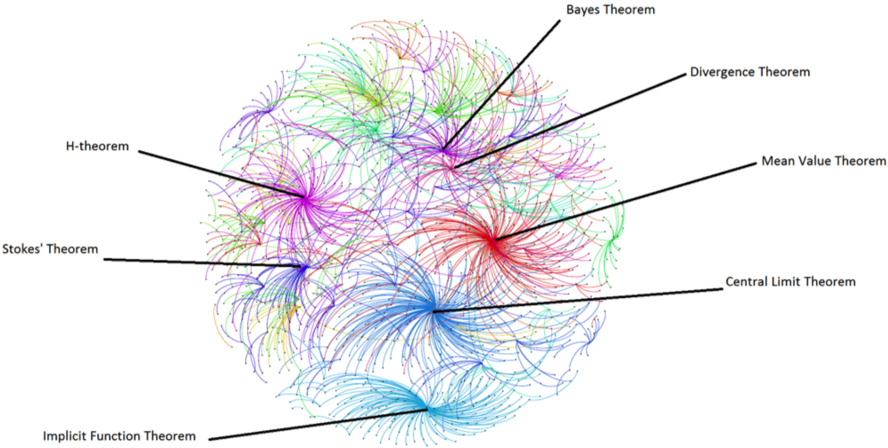
Community (4.#)	Count	Percentage
0	66	5.06%
1	18	1.38%
2	49	3.75%
3	9	0.69%
4	10	0.77%
5	17	1.30%
6	18	1.38%
7	48	3.68%
8	7	0.54%
9	4	0.31%
10	973	74.56%
11	15	1.15%
12	19	0.54%
13	4	0.31%
14	23	1.76%
15	19	1.46%
16	10	0.77%
17	2	0.15%
18	2	0.15%
19	2	0.15%
20	2	0.15%

Theorem ID	Theorem name	Frequency
9	fubini theorem	1008
319	sklar theorem	31
633	savitch theorem	6
820	perpendicular bisector theorem	14

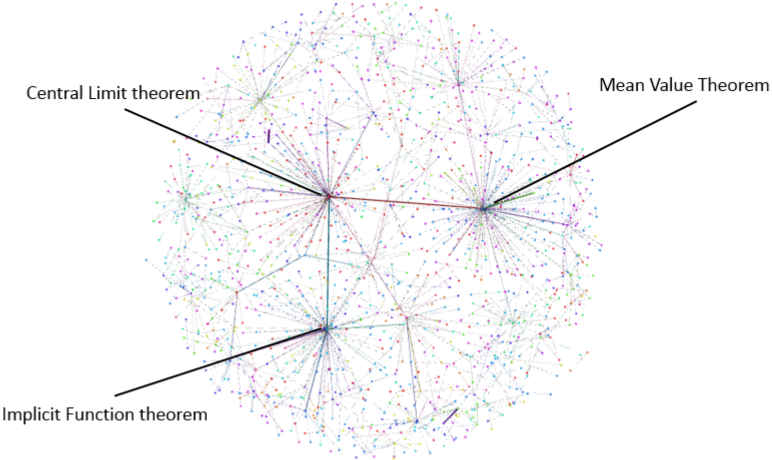
Inverse Laplacian Waiting Time Dynamics

Minimum spanning forest

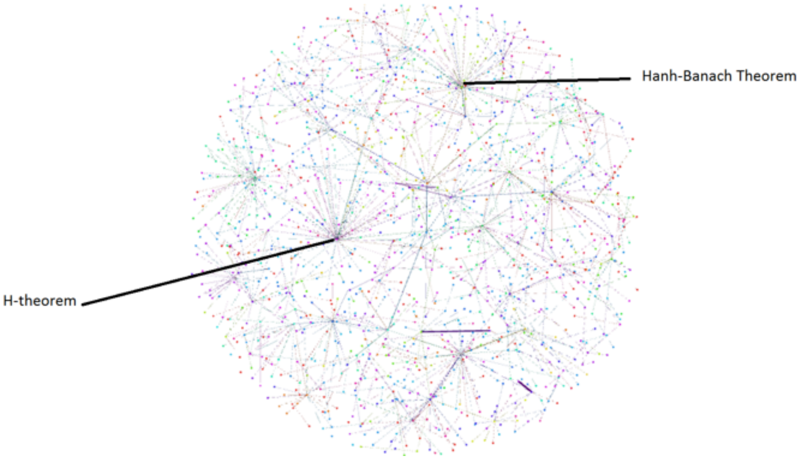
- A spanning forest of our network H includes all the nodes of H and contains no cycles
- We use Kruskal's algorithm to create a spanning forest



Maximum spanning forest



Maximum spanning forest



Maximum spanning forest