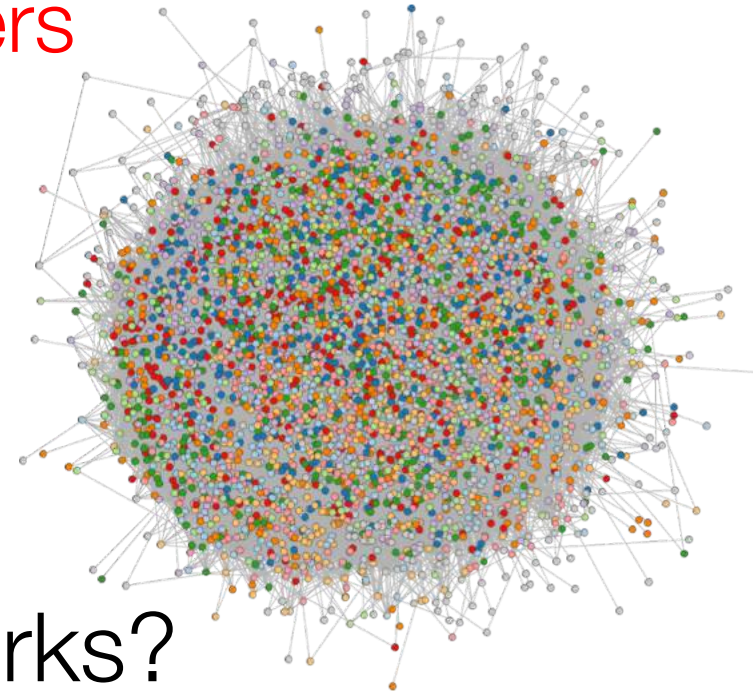


# **Large-Scale Structure of Networks**

# How do we characterize a network?

- Complex data: Hairball
  - Everything is inherently dependent to others
- What do they look like?
  - Statistical measures
    - Scalar value or distribution
- What is common in all real-world networks?
  - Recurring common patterns
  - Impacts how those networks work

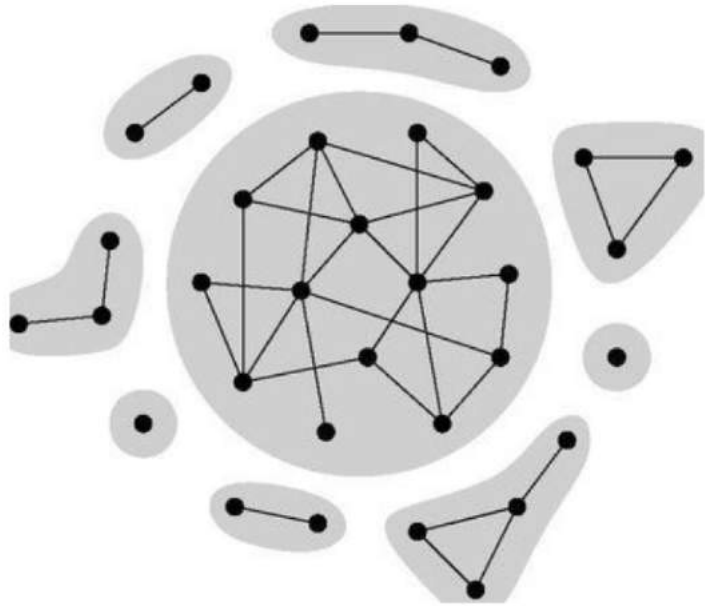


# Features

- Components
- Structural holes
- Homophily
- Structural balance
- Shortest paths & Small-world effect
- Degree distributions
- Power laws and scale-free networks
- Distribution of centrality metrics
- Clustering coefficient: Triangle closure
- Assortative mixing

# Components

- How many components?
  - What about their sizes?
- Giant connected component
  - And many small ones



	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	
	Physics coauthorship	Undirected	52 909	245 300	9.27	
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	
Technological	Internet	Undirected	10 697	31 992	5.98	
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	
	Physics coauthorship	Undirected	52 909	245 300	9.27	
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	
Technological	Internet	Undirected	10 697	31 992	5.98	
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	
Technological	Internet	Undirected	10 697	31 992	5.98	
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	
Technological	Internet	Undirected	10 697	31 992	5.98	
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967



# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	
	Peer-to-peer network	Undirected	880	1 296	1.47	
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	
	Protein interactions	Undirected	2 115	2 240	2.12	
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?

	Network	Type	<i>n</i>	<i>m</i>	<i>c</i>	<i>S</i>
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	
	Email address books	Directed	16 881	57 029	3.38	
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967



# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	
	Software classes	Directed	1 376	2 213	1.61	
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	
	Freshwater food web	Directed	92	997	10.84	
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
- Which ones?
  - Undirecteds?
  - Directeds?
    - Weakly conn. comp.

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	1.000
	Freshwater food web	Directed	92	997	10.84	1.000
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?
    - Directeds?
      - Weakly conn. comp.
- Can we have  $\geq 2$  large comps with similar sizes?
  - Graph theoretical property

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	1.000
	Freshwater food web	Directed	92	997	10.84	1.000
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?
    - Directeds?
      - Weakly conn. comp.
- Can we have  $\geq 2$  large comps with similar sizes?
  - Graph theoretical property
- No large, many smalls?

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	1.000
	Freshwater food web	Directed	92	997	10.84	1.000
	Neural network	Directed	307	2 359	7.68	0.967

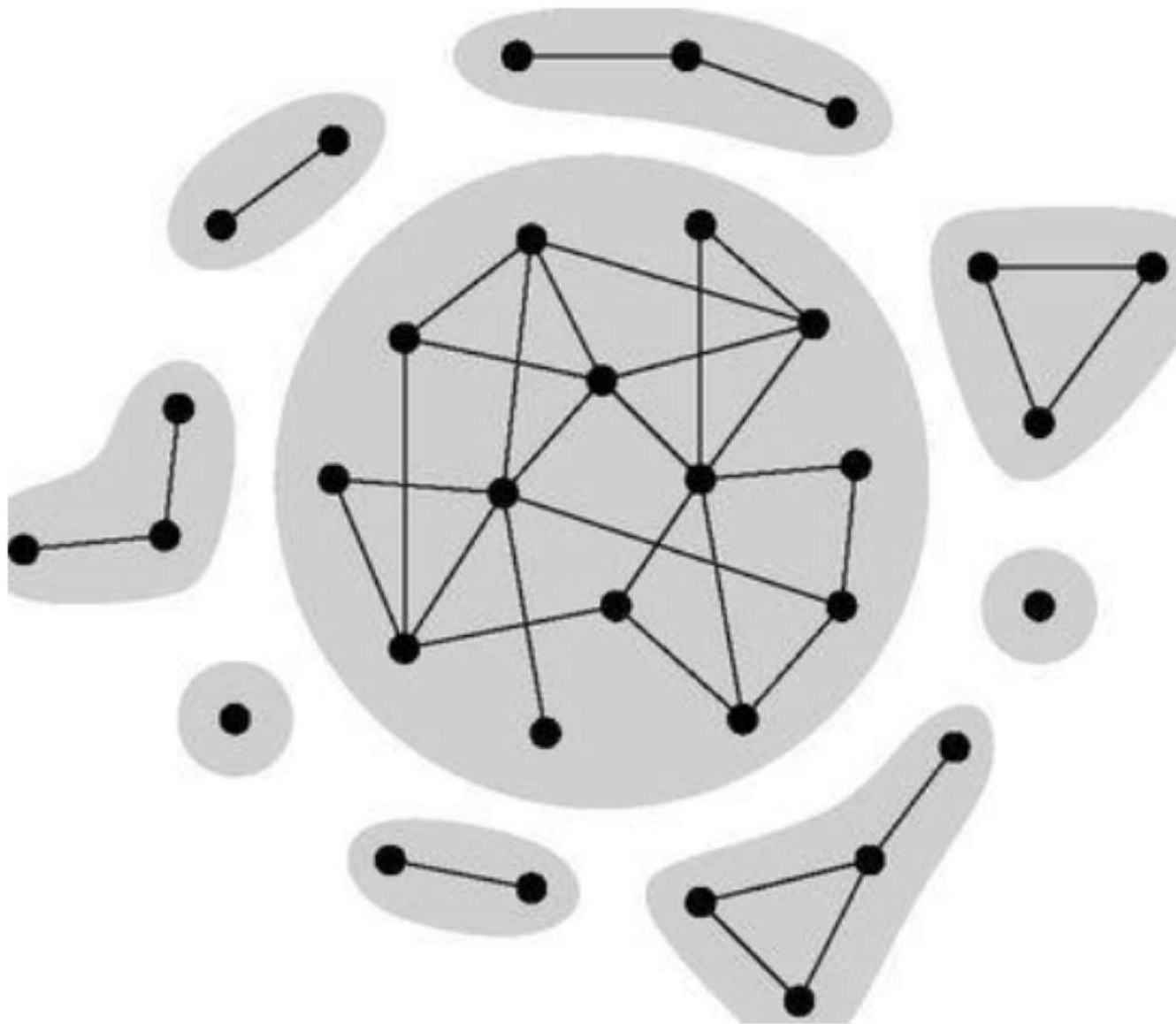
# Components

- Sometimes there is only one component
  - And there is a reason
  - Which ones?
    - Undirecteds?
    - Directeds?
      - Weakly conn. comp.
- Can we have  $\geq 2$  large comps with similar sizes?
  - Graph theoretical property
- No large, many smalls?
  - Well, we wouldn't care then

	Network	Type	$n$	$m$	$c$	$S$
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW nd . edu	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	1.000
	Freshwater food web	Directed	92	997	10.84	1.000
	Neural network	Directed	307	2 359	7.68	0.967

# Components

- Sometimes only one component
- And there are many types of components
  - Which one is better?
    - Undirected
    - Directed
      - Weakly
- Can we have  $\geq$  with similar size
  - Graph theoretic
- No large, many
  - Well, we would

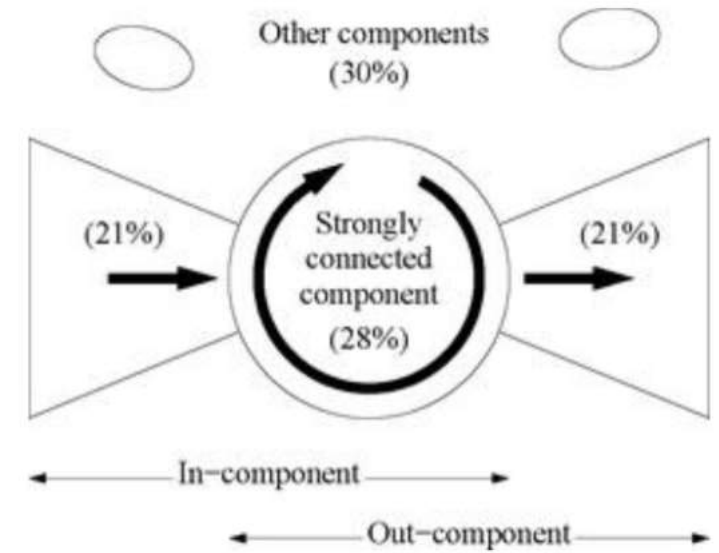


	<i>m</i>	<i>c</i>	<i>S</i>
	25 516 482	113.43	0.980
	55 392	14.44	0.876
	496 489	3.92	0.822
	245 300	9.27	0.838
	11 803 064	15.53	0.918
	80 000 000	3.16	
	86 300	1.44	0.952
	57 029	3.38	0.590
	477	1.66	0.503
	1 497 135	5.55	1.000
	1 466 000 000	7.20	0.914
	6 716 198	8.57	
	5 103	4.99	0.977
	16 100 000	66.96	1.000
	31 992	5.98	1.000
	6 594	2.67	1.000
	19 603	66.79	1.000
	1 723	1.20	0.998
	2 213	1.61	1.000
	53 248	4.34	1.000
	1 296	1.47	0.805
	3 686	9.64	0.996
	2 240	2.12	0.689
	598	4.46	1.000
	997	10.84	1.000
	2 359	7.68	0.967



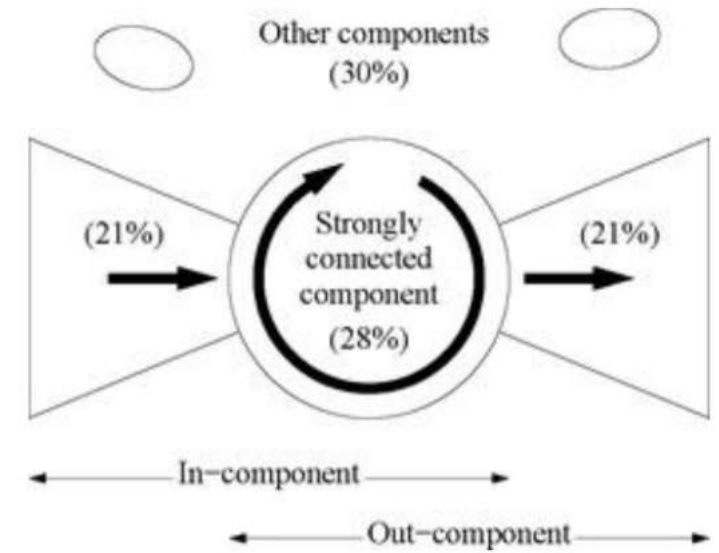
# Components

- Weakly connected component
  - One large, many small
- Strongly connected component?
  - Remember bow-tie?



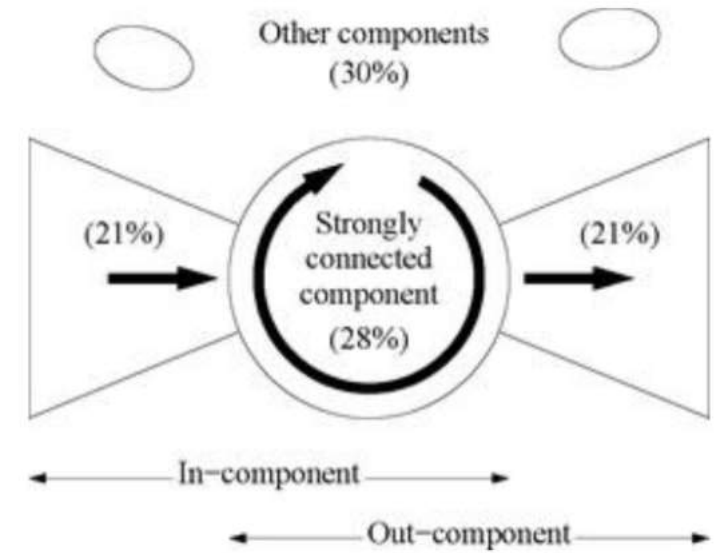
# Components

- Weakly connected component
  - One large, many small
- Strongly connected component?
  - Remember bow-tie?
- What about DAG?



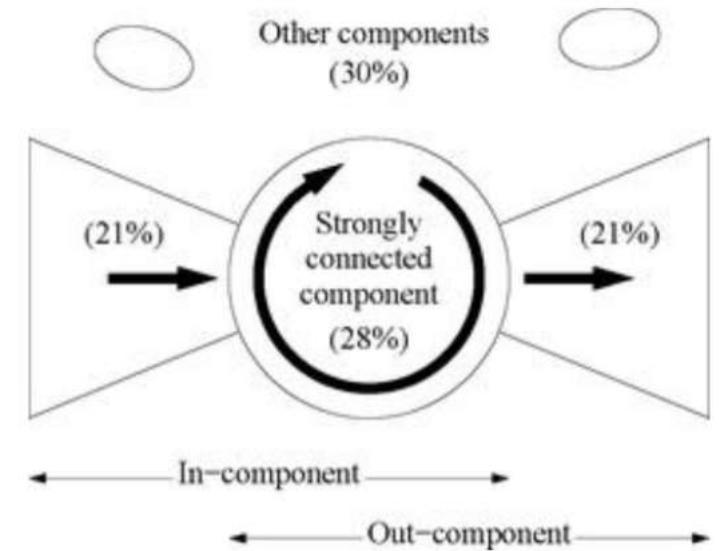
# Components

- Weakly connected component
  - One large, many small
- Strongly connected component?
  - Remember bow-tie?
- What about DAG?
  - No SCC with multiple nodes!
    - No cycle -> No SCC (of multiple nodes)



# Components

- Weakly connected component
  - One large, many small
- Strongly connected component?
  - Remember bow-tie?
- What about DAG?
  - No SCC with multiple nodes!
    - No cycle -> No SCC (of multiple nodes)
- How connected is the largest component in itself?

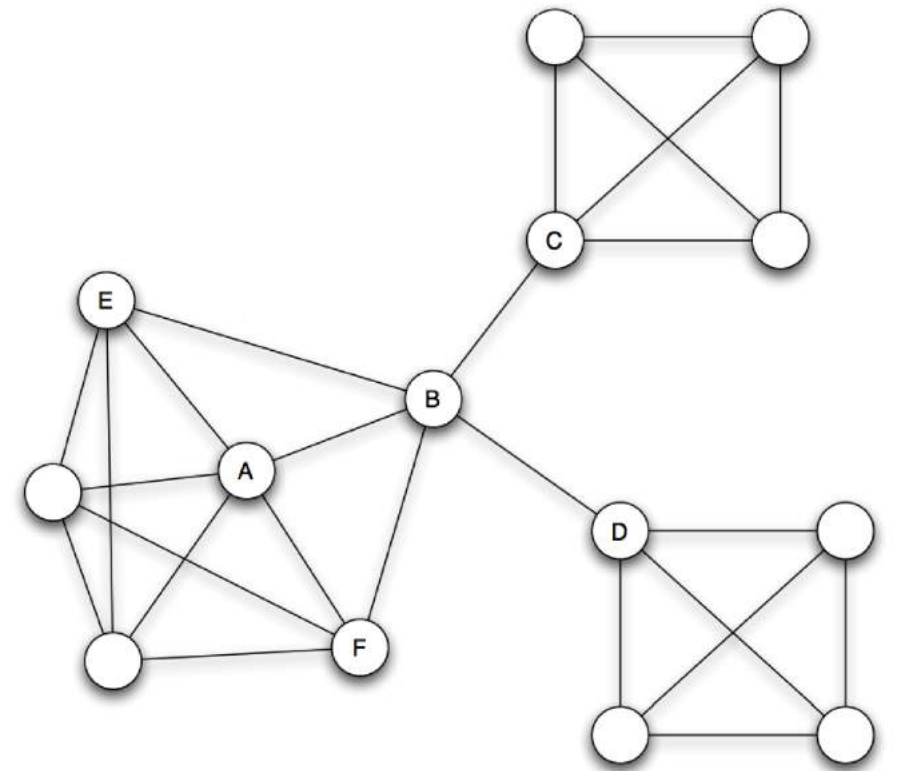


# Splitting Components

- How easily can you split a connected component?

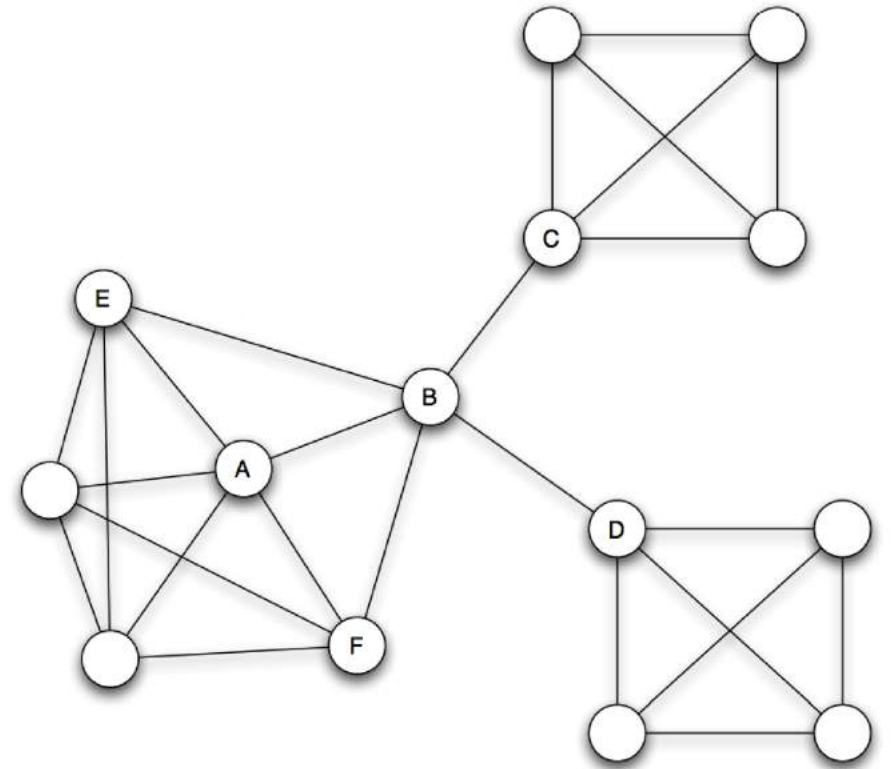
# Splitting Components

- How easily can you split a connected component?



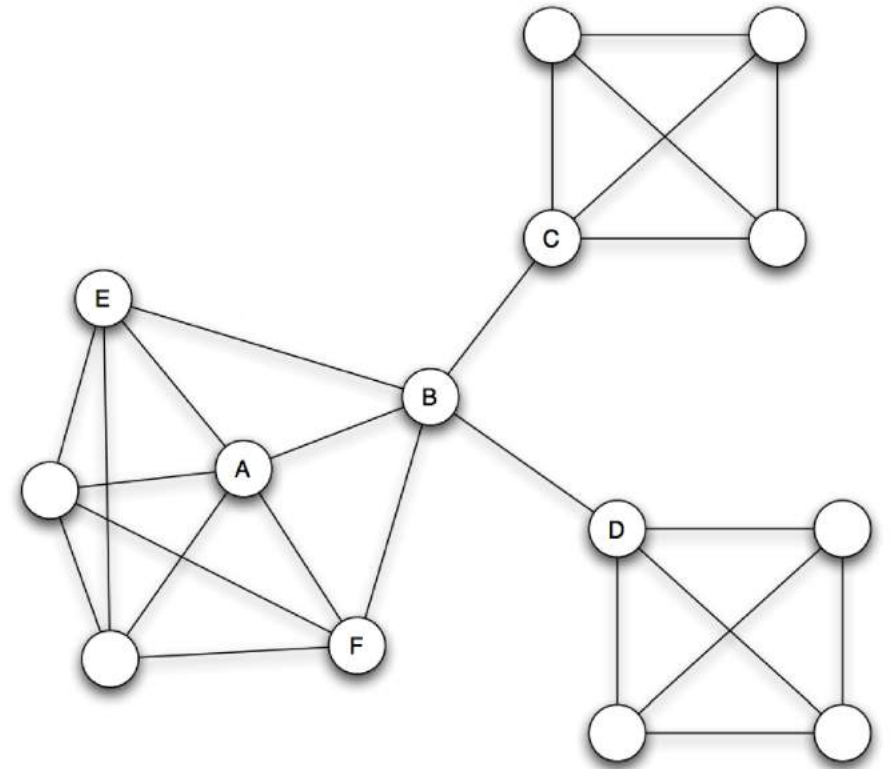
# Splitting Components

- How easily can you split a connected component?
  - Remove an edge
  - Remove a node
    - Removing multiple edges



# Splitting Components

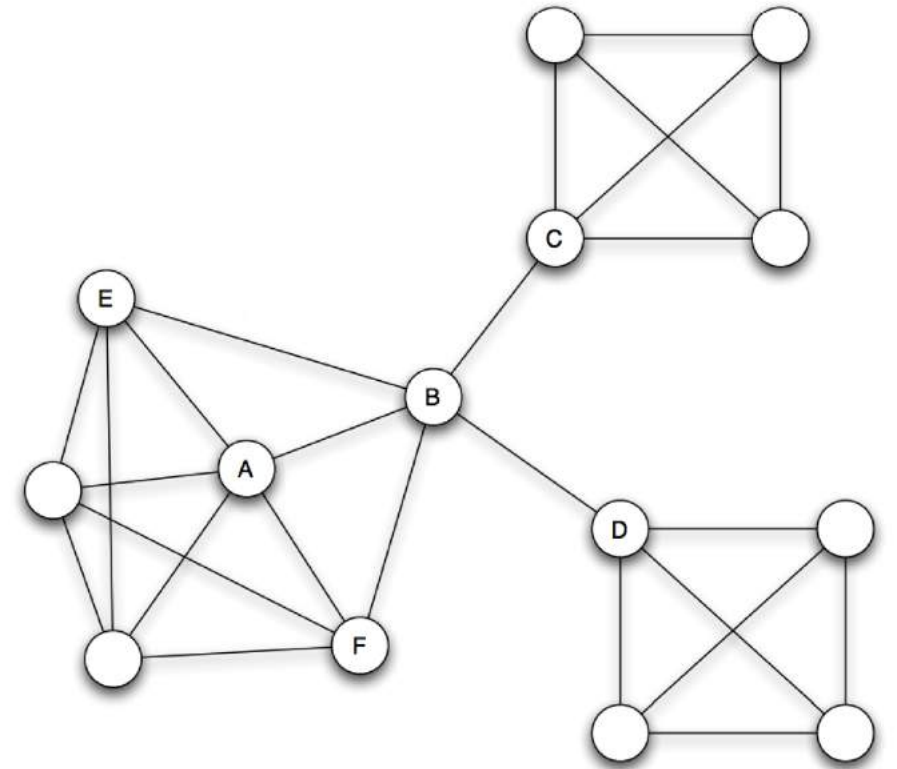
- How easily can you split a connected component?
  - Remove an edge
  - Remove a node
    - Removing multiple edges
- **Bridge**
  - Edge
  - Splits the graph if removed





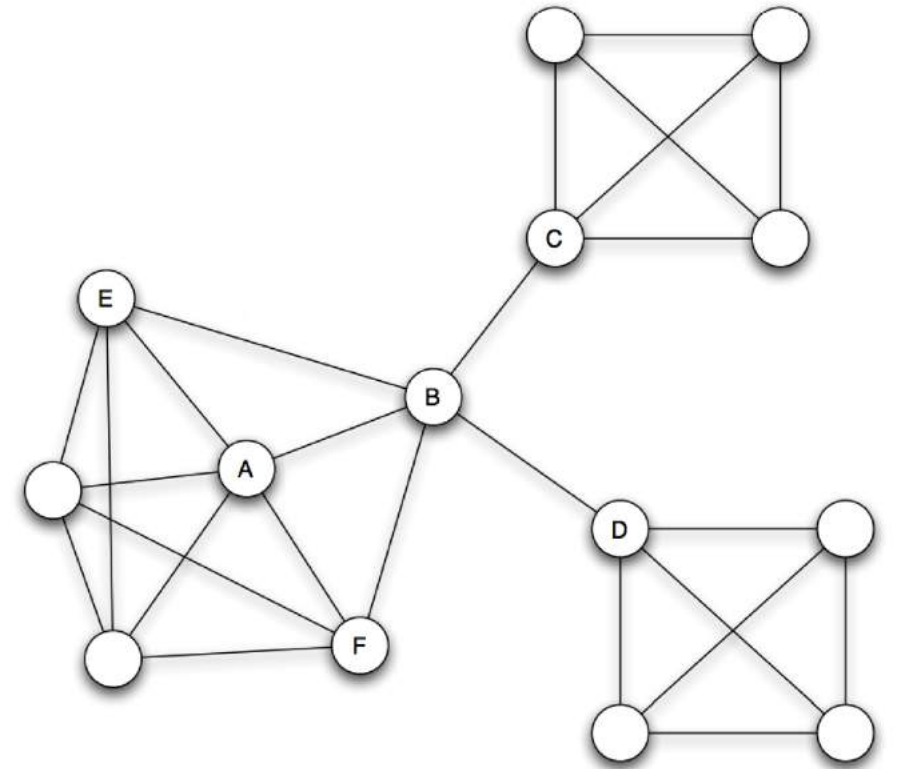
# Splitting Components

- How easily can you split a connected component?
  - Remove an edge
  - Remove a node
    - Removing multiple edges
- **Bridge**
  - Edge
  - Splits the graph if removed
- **Articulation point**
  - Node
  - Splits the graph if removed



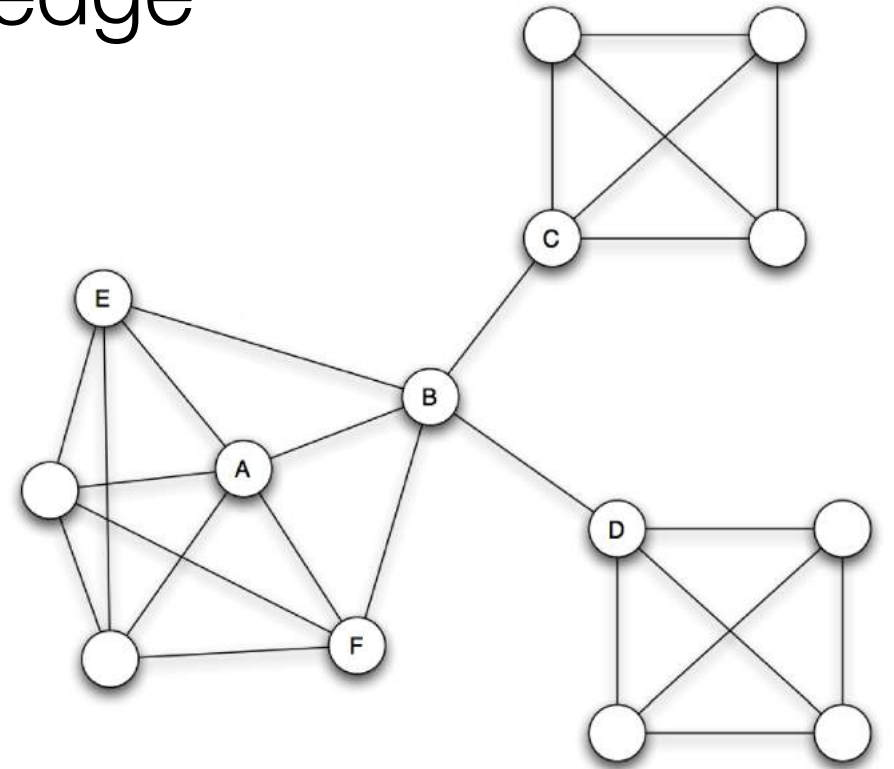
# Roles of the nodes and edges

- Not all nodes/edges have the same role



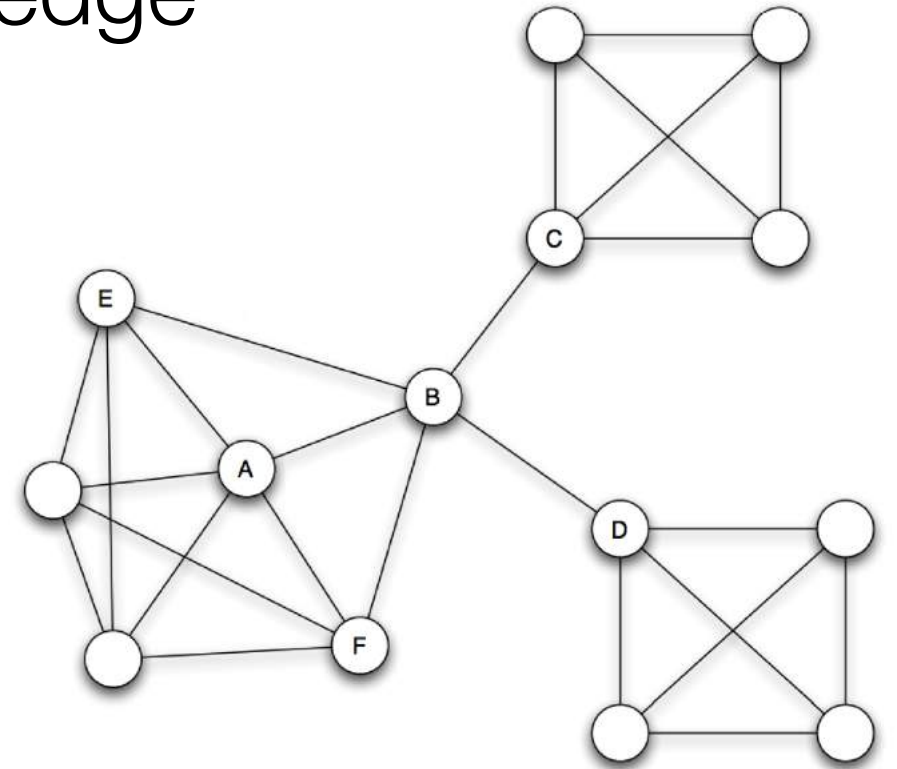
# Roles of the nodes and edges

- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points



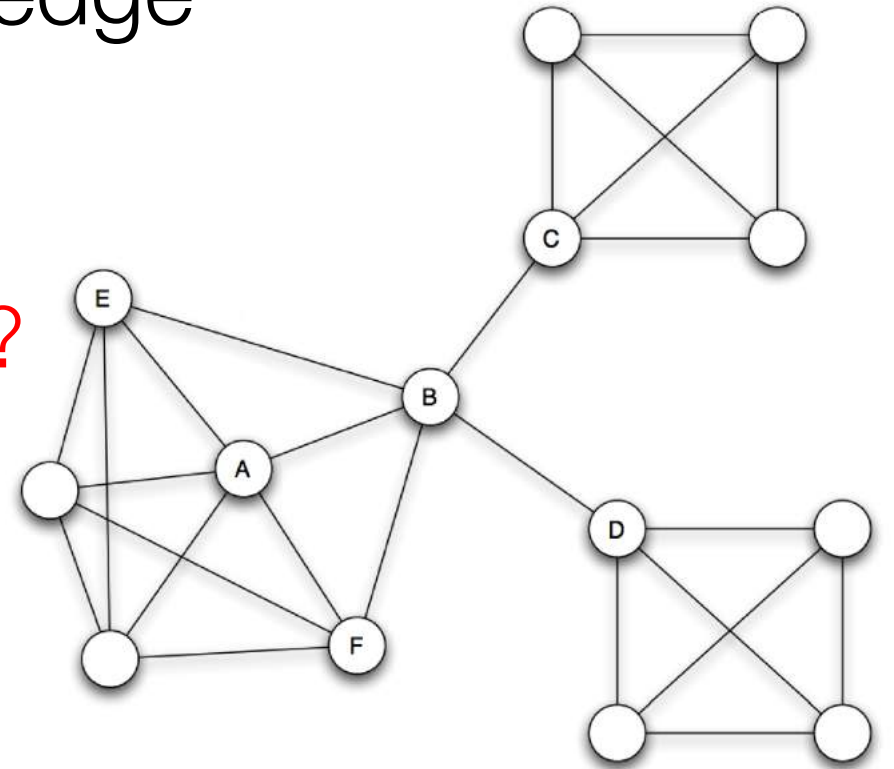
# Roles of the nodes and edges

- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points
- Bridge has 0 support



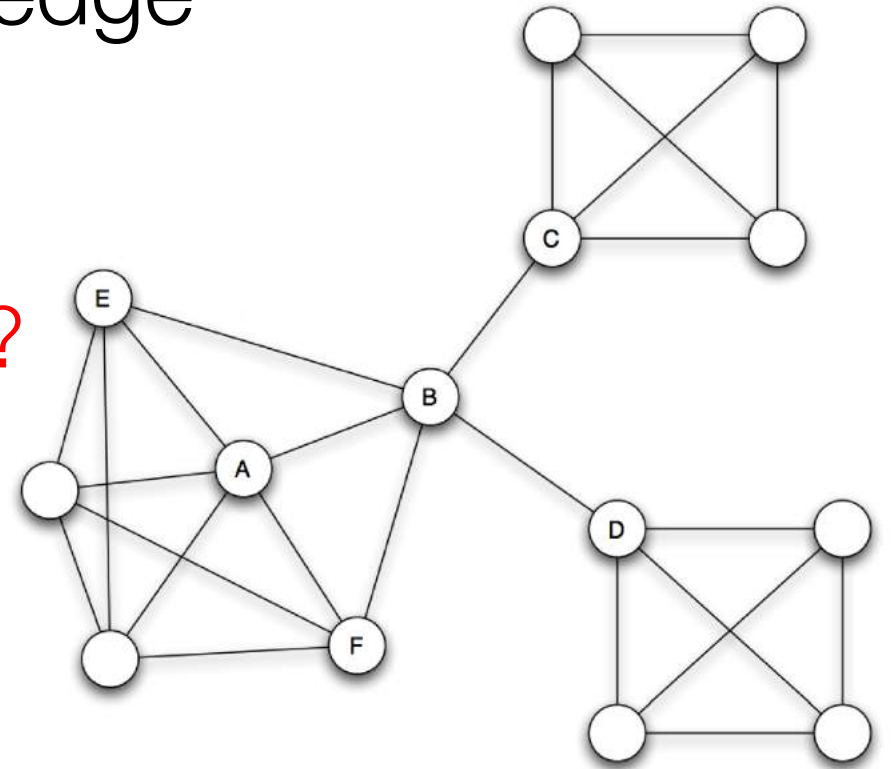
# Roles of the nodes and edges

- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points
- Bridge has 0 support
  - An edge with no support is a bridge?



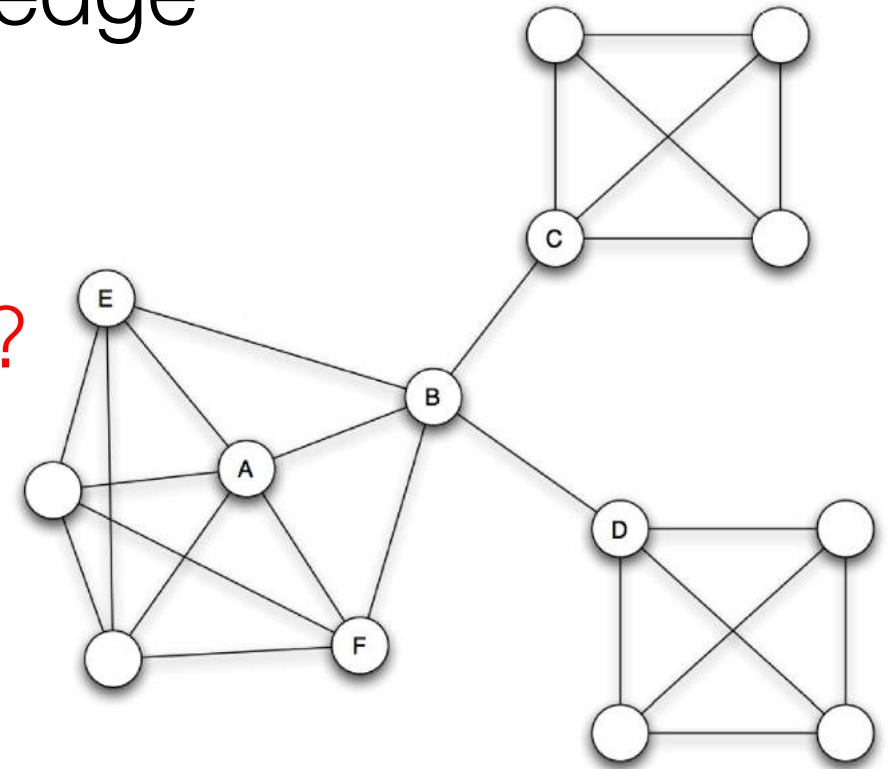
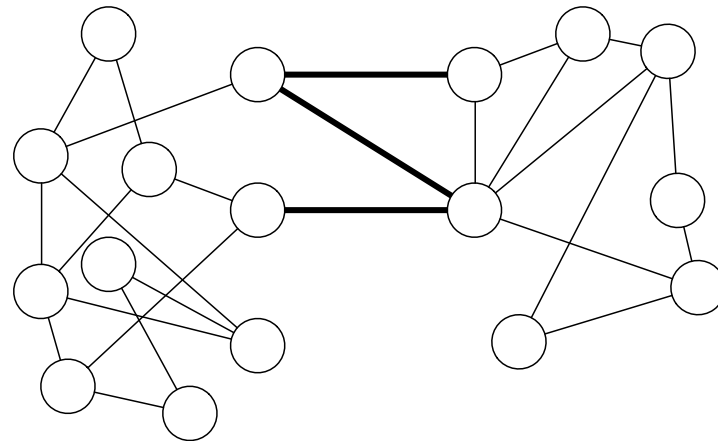
# Roles of the nodes and edges

- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points
- Bridge has 0 support
  - An edge with no support is a bridge?
  - NO



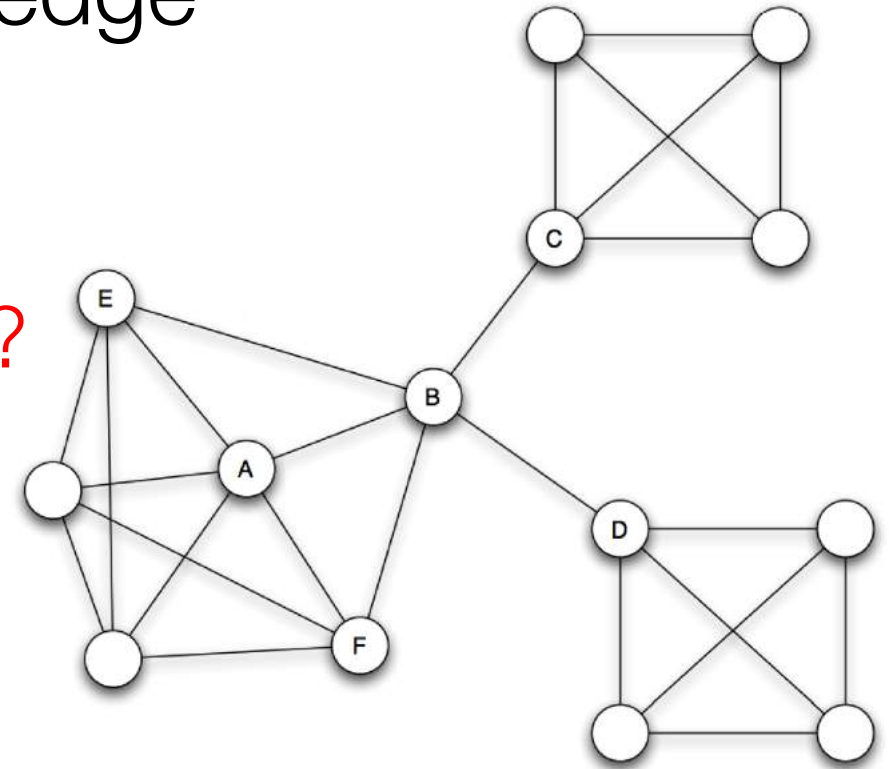
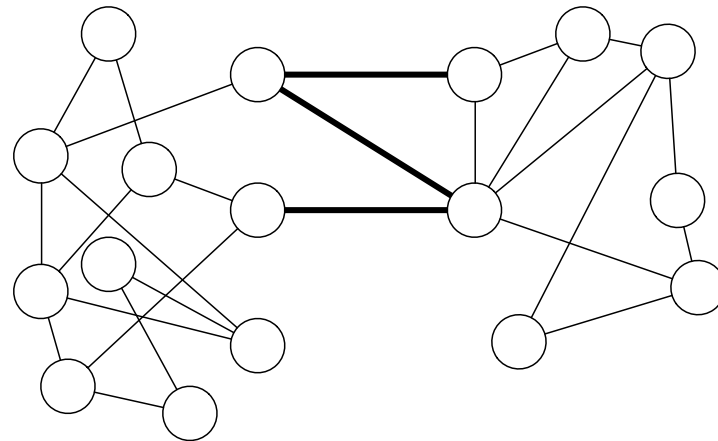
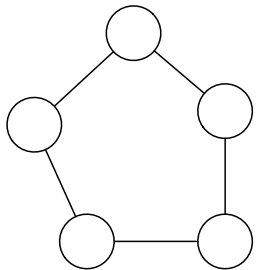
# Roles of the nodes and edges

- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points
- Bridge has 0 support
  - An edge with no support is a bridge?
    - NO
    - Local bridge



# Roles of the nodes and edges

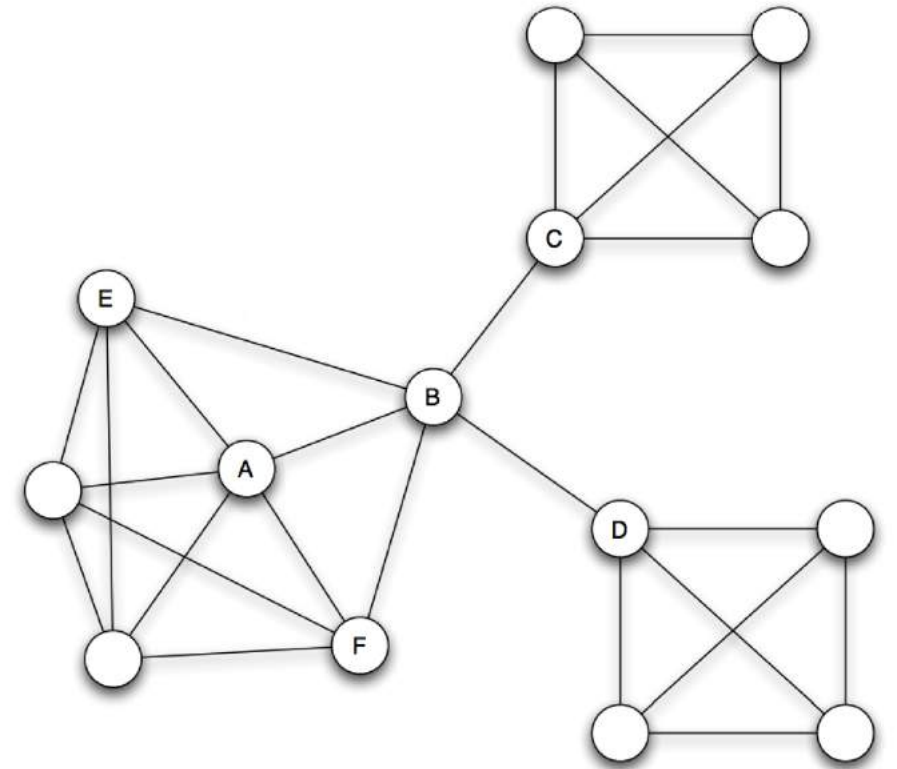
- Not all nodes/edges have the same role
- Embeddedness (or support) of an edge
  - # common neighbors of end points
- Bridge has 0 support
  - An edge with no support is a bridge?
    - NO
    - Local bridge
    - 3+ cycles





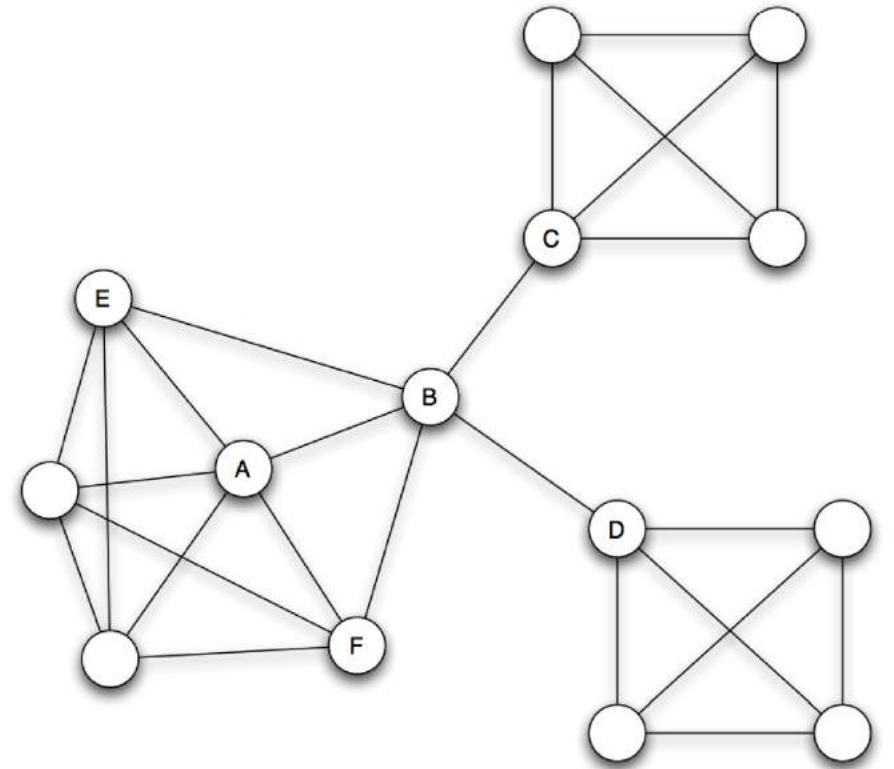
# Roles of the nodes and edges

- Which one is more advantageous?
  - A or B?



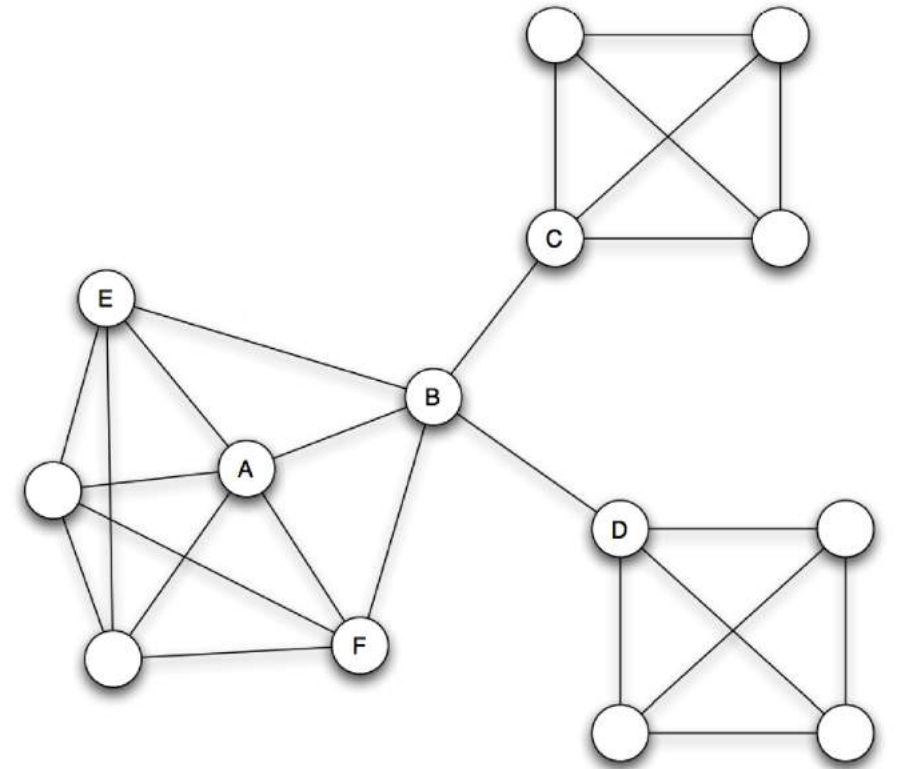
# Roles of the nodes and edges

- Which one is more advantageous?
  - A or B?
- Relations of A are more trusty
  - Cannot be said for B



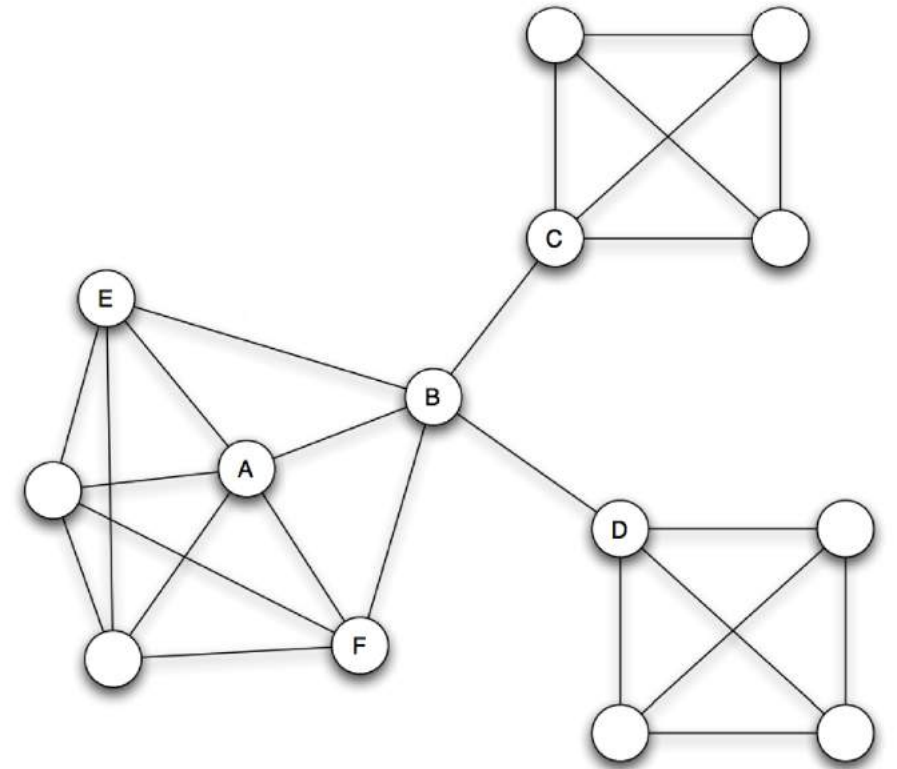
# Roles of the nodes and edges

- Which one is more advantageous?
  - A or B?
- Relations of A are more trusty
  - Cannot be said for B
- B has early access to information



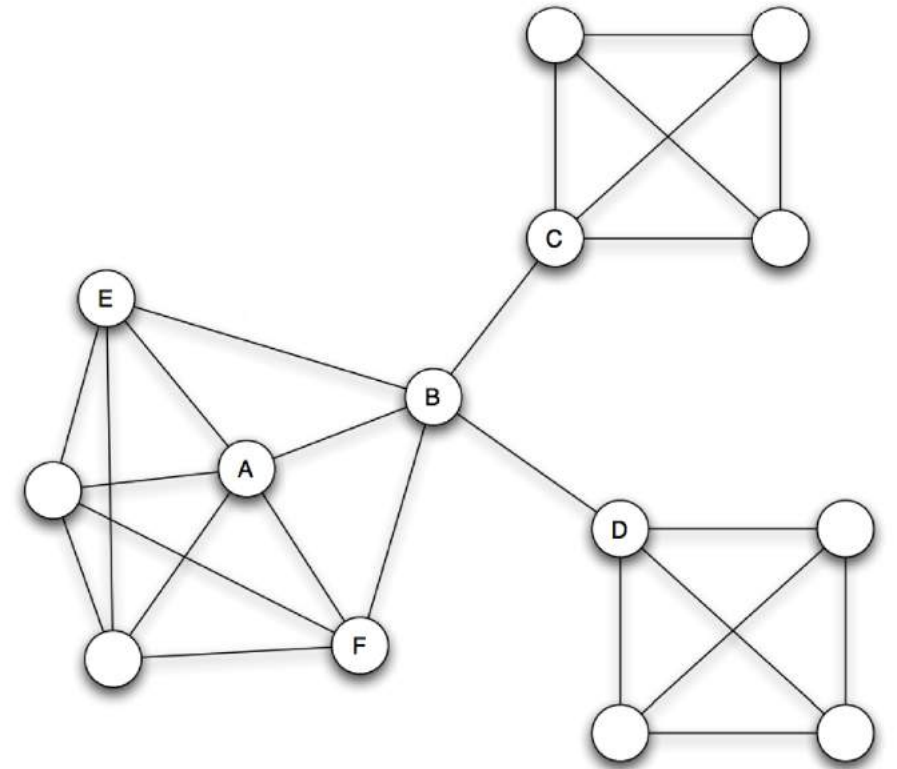
# Roles of the nodes and edges

- Which one is more advantageous?
  - A or B?
- Relations of A are more trusty
  - Cannot be said for B
- B has early access to information
  - Also likely to invent first
  - And also a gatekeeper
    - Controls the information flow
  - A is not like that



# Roles of the nodes and edges

- Which one is more advantageous?
  - A or B?
- Relations of A are more trusty
  - Cannot be said for B
- B has early access to information
  - Also likely to invent first
  - And also a gatekeeper
    - Controls the information flow
  - A is not like that
- B is called a structural hole



# Homophily

- Birds of a feather flock together
  - We tend to be similar to our friends



# Homophily

- Birds of a feather flock together
  - We tend to be similar to our friends



- How to consider this in the network context?

# Homophily

- Birds of a feather flock together
  - We tend to be similar to our friends
- How to consider this in the network context?
  - Your friends are not a random sample





# Homophily

- Birds of a feather flock together
  - We tend to be similar to our friends
- How to consider this in the network context?
  - Your friends are not a random sample
  - Similar age
  - Racial, ethnic dimensions



# Homophily

- Birds of a feather flock together
  - We tend to be similar to our friends
- How to consider this in the network context?
  - Your friends are not a random sample
  - Similar age
  - Racial, ethnic dimensions
  - Less-mutable characteristics
    - Place you live, jobs you do, beliefs, opinions
  - Of course not all friends are like that, but this is usual

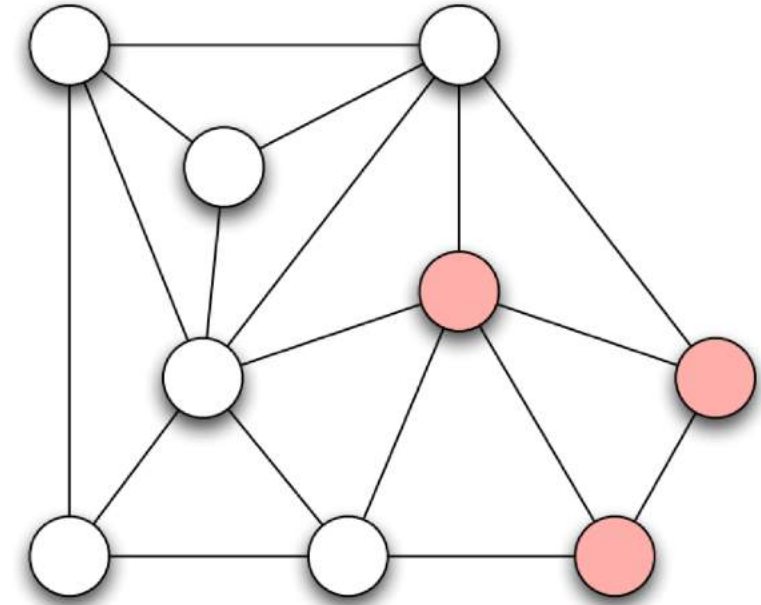


# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!

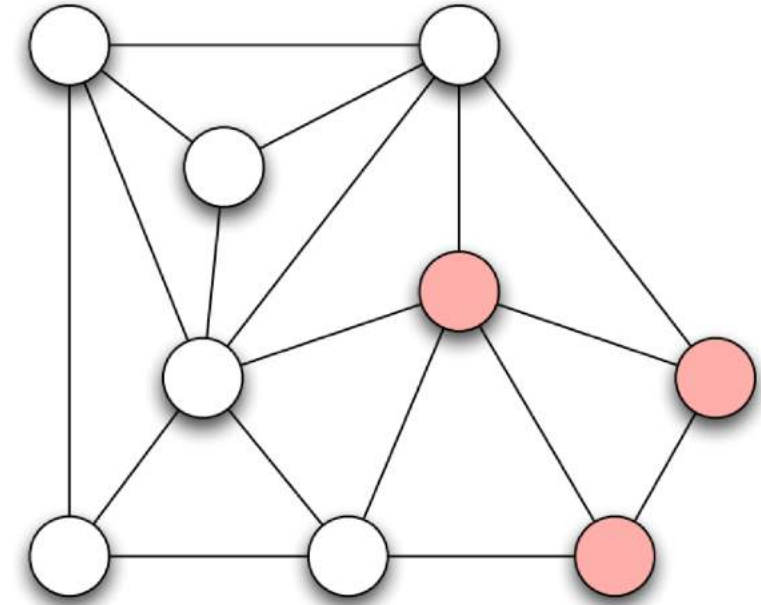
# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders



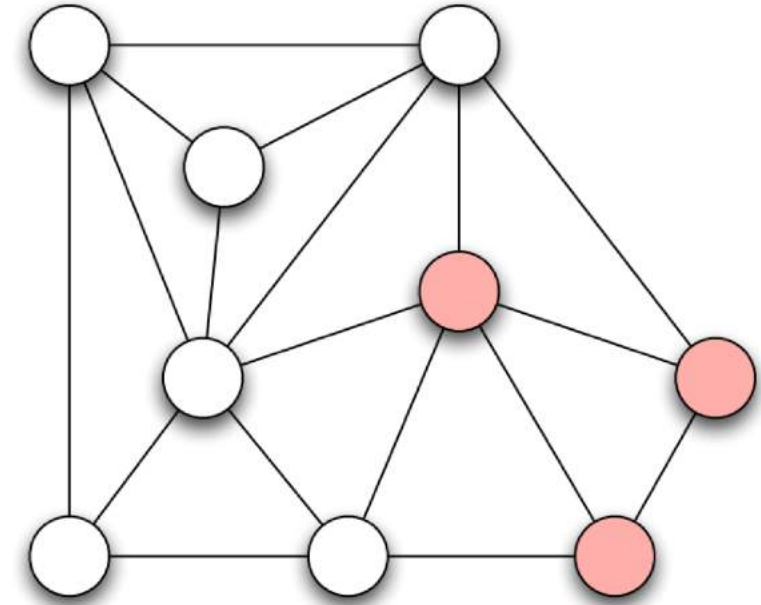
# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders
  - How to measure the homophily?



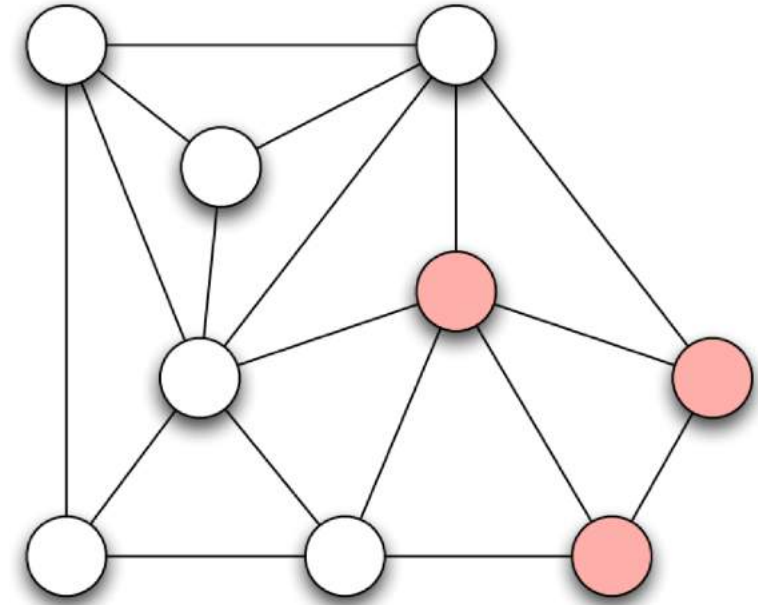
# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders
  - How to measure the homophily?
  - $p$  males,  $q$  female (fractions)
  - Let's randomly assign genders
    - $p^2$  male edges,  $q^2$  female edges
    - $2pq$  male-female edges



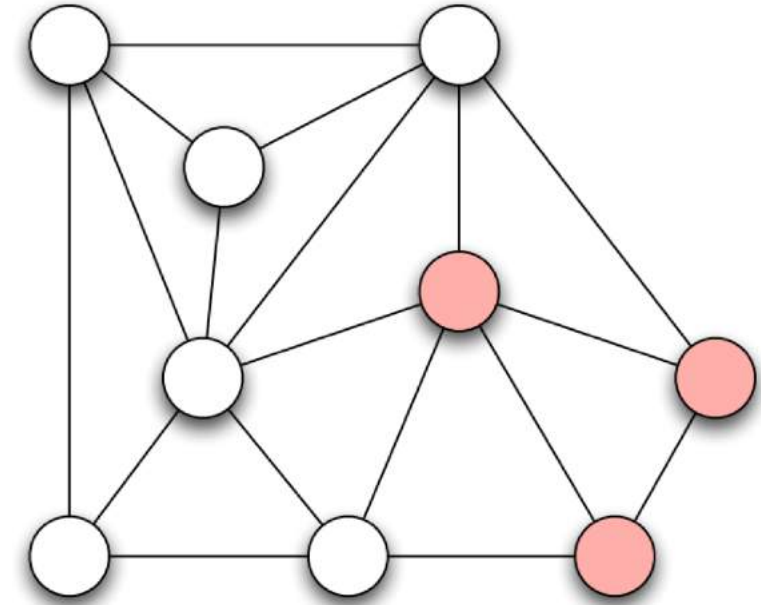
# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders
  - How to measure the homophily?
  - $p$  males,  $q$  female (fractions)
  - Let's randomly assign genders
    - $p^2$  male edges,  $q^2$  female edges
    - $2pq$  male-female edges
      - If real fraction of such edges is significantly less; then homophily exists



# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders
  - How to measure the homophily?
  - $p$  males,  $q$  female (fractions)
  - Let's randomly assign genders
    - $p^2$  male edges,  $q^2$  female edges
    - $2pq$  male-female edges
      - If real fraction of such edges is significantly less; then homophily exists
  - What if the real value is significantly more than  $2pq$ ?





# Features of nodes

- Degrees
  - High-degree nodes are connected; assortativity!
- Genders
  - How to measure the homophily?
  - $p$  males,  $q$  female (fractions)
  - Let's randomly assign genders
    - $p^2$  male edges,  $q^2$  female edges
    - $2pq$  male-female edges
      - If real fraction of such edges is significantly less; then homophily exists
  - What if the real value is significantly more than  $2pq$ ?
    - Inverse homophily -> romantic relationships

