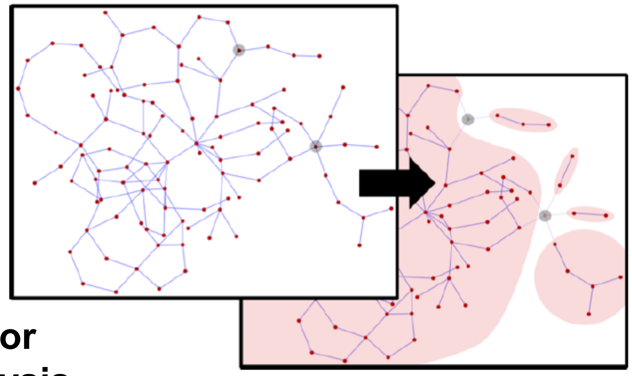


Graph Theory for Structural Analysis

TTM4185 SIKKERHET og ROBUSTHET i IKT System

Yuming Jiang

(Slides are based on an earlier version prepared by Dr. Jonas Wäfler @ Powel)

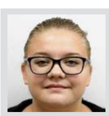
NTNU Norwegian University of Science and Technology

Who is him?

- Professor at IIK (and former ITEM), since 2005.
- TTM4185 course responsible since 2019.
- Research areas:
  – **Network Performance, Dependability and Security**

Karin Bernsmed
Lecturer

email: karin.bernsmed@ntnu.no
subject field [ttm4185]



Lise Millerjord
Lecturer, exercises

room: B-217
email: lise.millerjord@ntnu.no
subject field [ttm4185]

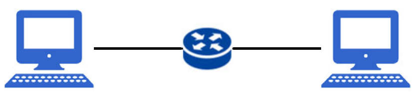
Yuming Jiang
Lecturer, course responsible

email: yuming.jiang@ntnu.no
subject field [ttm4185]

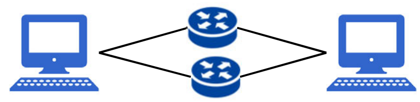
NTNU Norwegian University of Science and Technology

Which is more robust / dependable?

Network A

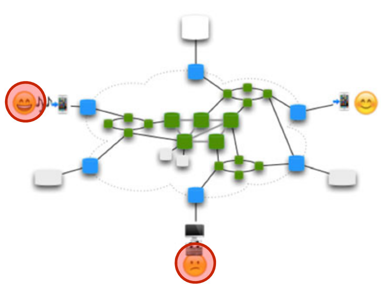


Network B

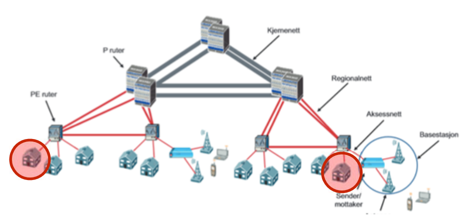


Which is more robust / dependable?

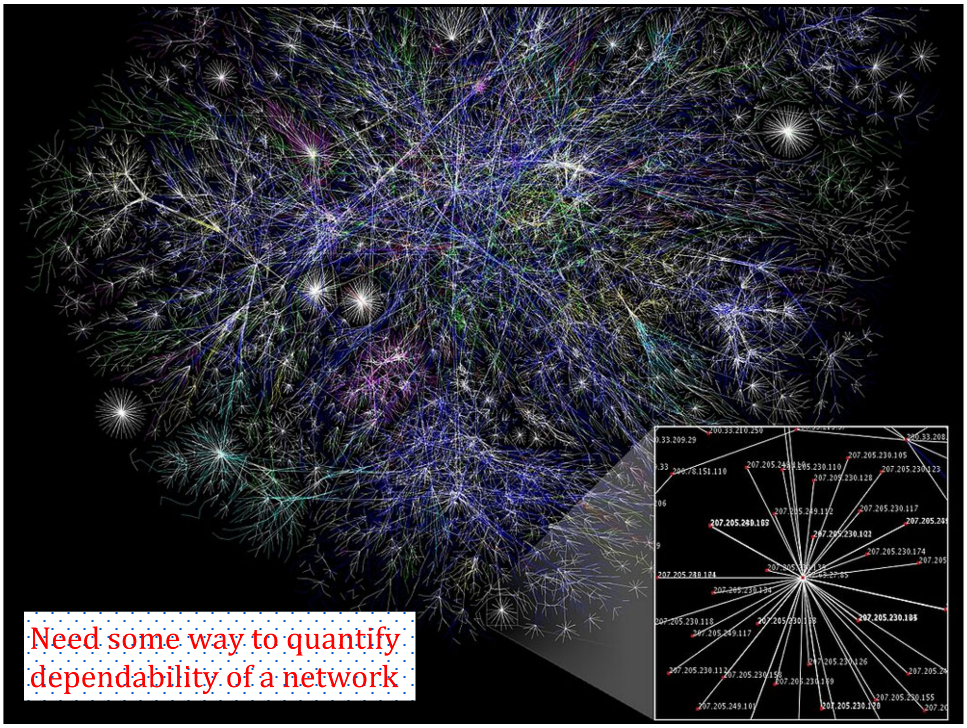
Network A



Network B



- *Difficult to tell from immediate impression*



Why structural analysis with graphs?

- Analyze and understand networks
  - Dependencies and structures
  - Fault-tolerance
  - Attack resistance
  - Traffic flow and routing
- Model networks and network services as graphs:
  - Physical networks
  - Network of virtual connections
  - Chain of sub-services / functions

NTNU Norwegian University of Science and Technology

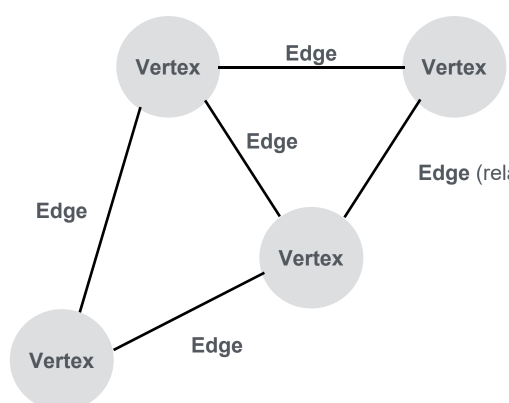
Agenda

- About Graphs (Repetition)

Structural Analysis of graphs/networks

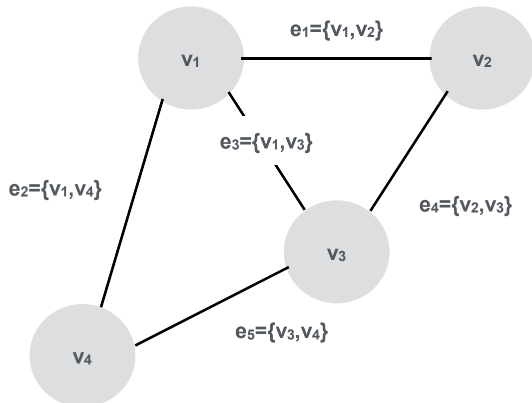
- Network Properties
  - Basic Network Properties
  - Node Degree (and Random Networks)
  - Centrality
  - Flow Analysis
- Dependability analysis in networks
  - Robustness

About Graphs - Basic Concepts



Edge (relation between vertices)

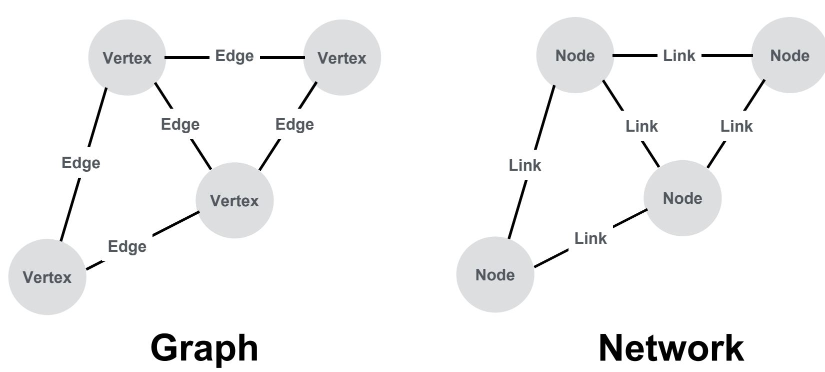
About Graphs - Basic Concepts



$G = \{ V, E \} = \{ \{v_1,\ldots,v_4\}, \{e_1,\ldots,e_5\} \}$

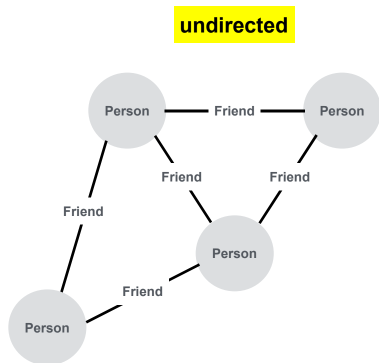
NTNU Norwegian University of Science and Technology

About Graphs - Graph vs. Network

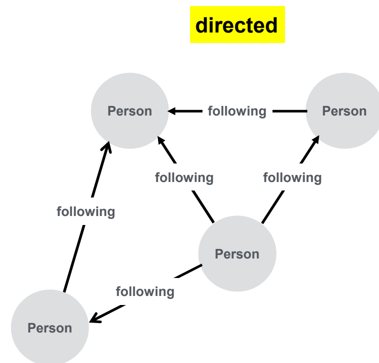


NTNU Norwegian University of Science and Technology

About Graphs - Network example



Social Network



Social Network

About Graphs
Why Graph Theory/Network Science?

1. Convert problem into graph
2. Use tools from graph theory to get results
3. Interpret graph theory result in original setting

→ Real systems of different nature can have the same graph representation

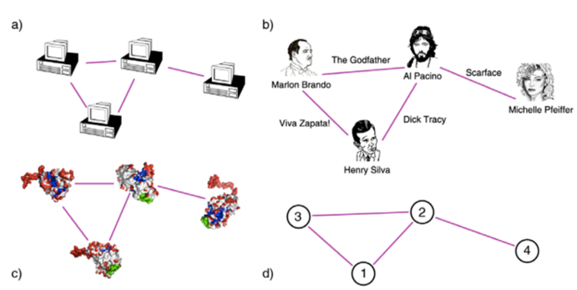


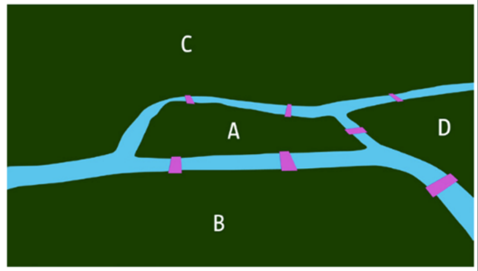
Figure: ***Network Science*** by Albert-László Barabási

About Graphs
Example 1: Bridges of Königsberg

- "Can one have a walk across all seven bridges and never cross the same bridge twice?"
- 1735: Leonard Euler found mathematical proof using graph theory

1. Convert problem to graph
2. Use graph theory
   1. Nodes with odd number of links have to be start or end point of path
   2. All 4 nodes have odd node degree → not possible
3. Walk is not possible!

(Now known as Euler/Eulerian path: Euler's Theorem - A connected graph has a Euler cycle/path if and only if every vertex has even degree.)



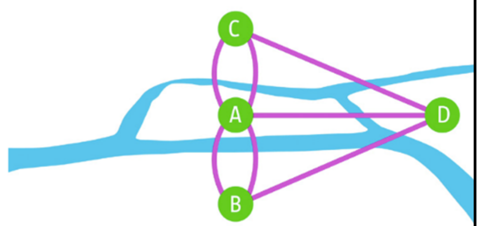
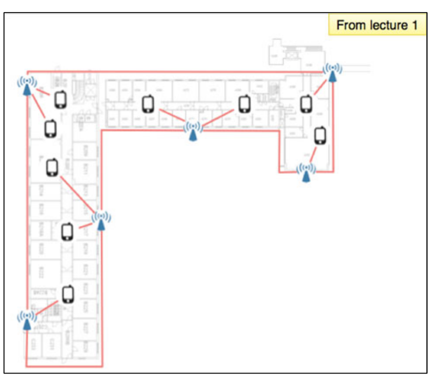


Figure: ***Network Science*** by Albert-László Barabási

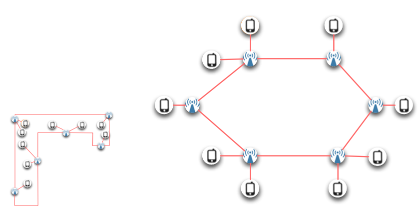
About Graphs
Example 2: communication network



Example

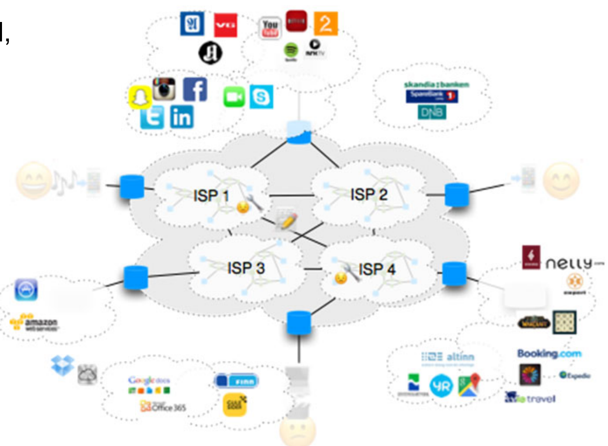
- Routing between mobile devices

1. Model network as graph:
   – Vertices: mobile devices, access points
   – Edges: wireless / wired links
2. Find shortest path in graph
3. Shortest routing between devices



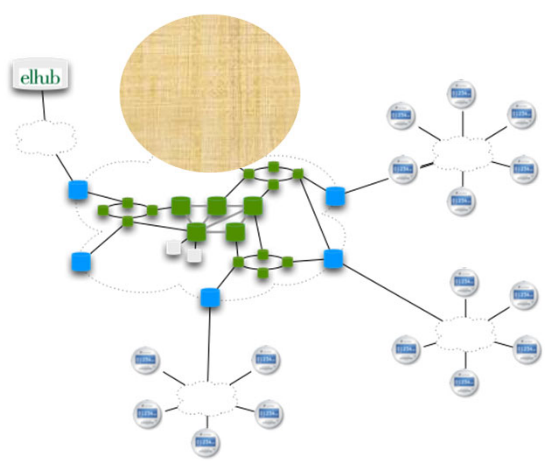
Service (Chain)

- A service chain consists of functional, physical and virtual components, which form a structure (topology).
- *Insight into the full topology is not possible…*



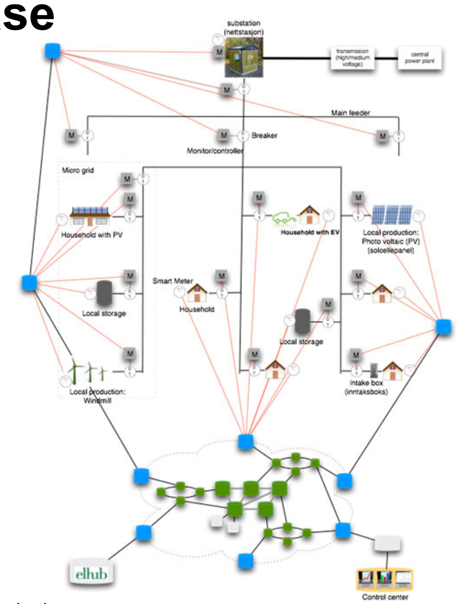
The smart grid case

- *For a single service, provided by one operator, it might be possible* to make a mapping of the topology. However, even then we do not necessarily have full insight.



The smart grid case

- For a single service, provided by one operator, it might be possible to make a mapping of the topology. But, even then we do not necessarily have full insight.



Modelling

There is no single solution in modelling!

Abstraction levels
- Physical connections
- Slightly abstracted
- Logical connections
  - i.e. on higher network layer

https://en.wikipedia.org/wiki/Network_theory

Agenda

- About Graphs (Repetition)

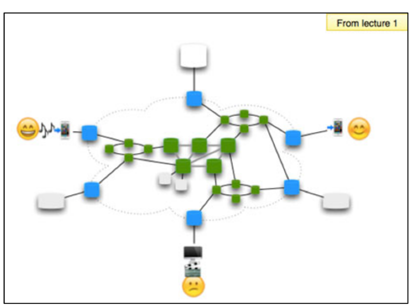
Structural Analysis of graphs/networks

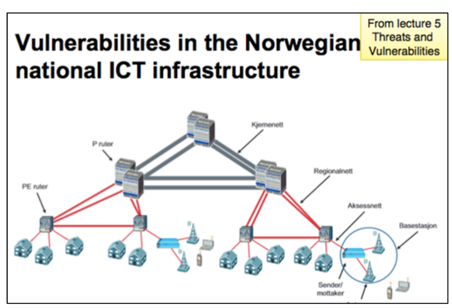
- Network Properties
  - Basic Network Properties
  - Node Degree (and Random Networks)
  - Centrality
  - Flow Analysis
- Dependability analysis in networks
  - Robustness

NTNU Norwegian University of Science and Technology

Structural analysis Network properties

- How to compare the structure of networks?
- How to assess its stability/robustness/dependability?
- Which nodes are more important/central in network?
- Which nodes need to have a higher level of dependability?





NTNU Norwegian University of Science and Technology

© Course TTM4185, Department of Information Security and Communication Technology , NTNU

Network properties
Basic properties

- **Number of vertices**:
  - VertexCount[GraphA]= 4
  - VertexCount[GraphB]= 12
- **Number of edges**:
  - EdgeCount[GraphA]=4
  - EdgeCount[GraphB]=12
- **Graph Diameter**:
  "longest shortest path in a graph"
  (longest distance between any node pair)
- **Connectivity**:
  Graph is *connected* if there is a path between any pair of vertices in the graph

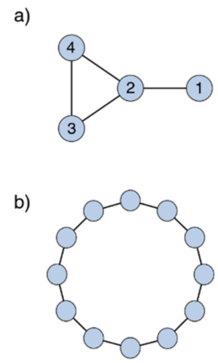


Figure: ***Network Science*** by Albert-László Barabási

Network properties
Vertex Degree

- **Vertex Degree**
  Number of links of the vertex
  - VertexDegree[GraphA]= {1,3,2,2}
  - VertexDegree[GraphB]= {2,2,2,2,2,2,2,2,2,2,2,2}
- **Average Vertex Degree**
  - Mean[VertexDegree[GraphA]]= 2
  - Mean[VertexDegree[GraphB]]= 2

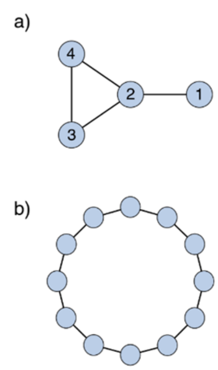


Figure: ***Network Science*** by Albert-László Barabási

Network properties
Vertex/Node Degree

- **Vertex Degree**
  Number of links per node
  – VertexDegree[GraphA]= {1,3,2,2}
  – VertexDegree[GraphB]= {2,2,2,2,2,2,2,2,2,2,2,2}
- **Average Vertex Degree**
  – Mean[VertexDegree[GraphA]]= 2
  – Mean[VertexDegree[GraphB]]= 2
- **(Node) Degree Distribution**
  distribution denoting probability that a node has *k* nodes

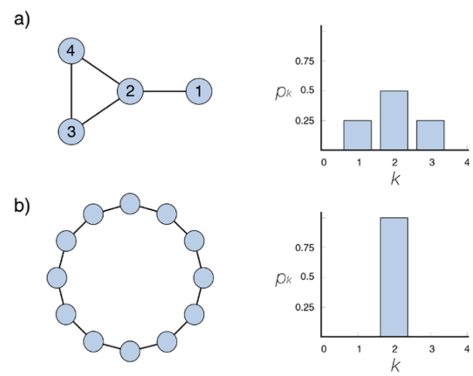


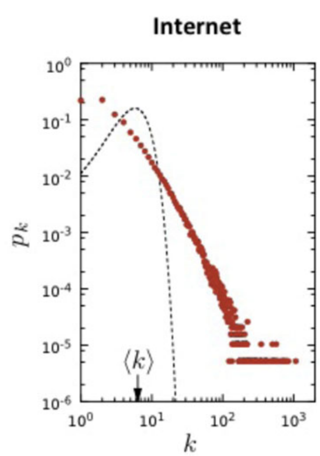
Figure: ***Network Science*** by Albert-László Barabási

Vertex/Node Degree Distribution
Example: Internet

- Degree distribution of the Internet
  – *Power law* distribution $p_k \sim k^{-\gamma}$
  – Linear in log-log plot
  – Also called scale-free networks

- Scale-free networks occur in many areas of science and engineering:
  – Power grid
  – Social networks
  – Etc.

- Many nodes with few connections, few nodes (hubs) with many connections



Connectivity after node failure?

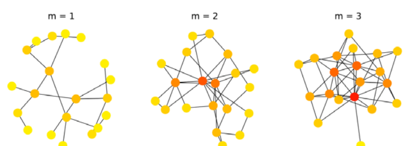
Figure: ***Network Science*** by Albert-László Barabási

Vertex/Node Degree Distribution
Random Networks

- Degree distribution can be used to create random networks with desired properties

Models

- Erdős–Rényi
- Barabási-Albert
- Watts-Strogatz



Source: wikipedia

Barabási-Albert algorithm

- Creates random network with power law distribution
  $p_k \sim k^{-\gamma}$
- Based on preferential attachment
  - Start with *m* connected nodes
  - Add new node and connect to *m* existing nodes
  - Connection to new nodes with probability

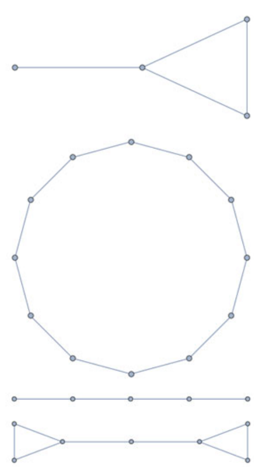
$$p_i = k_i / \sum_j k_j$$

"depending on fraction of total links connected to node i"

Network properties
Centrality

Different centrality measures:

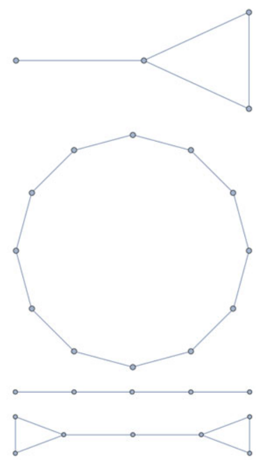
- Which node is the most important/central one in each network?



Network properties
Centrality

Different centrality measures:

- **Degree Centrality**
  node degree = centrality index
  "node with most neighbors is most central"

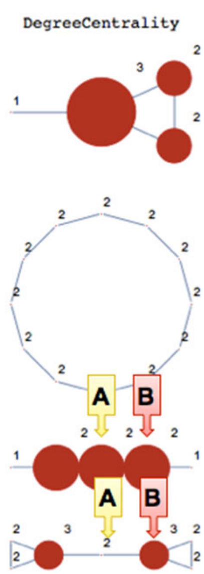


Network properties
Centrality

Different centrality measures:

- **Degree Centrality**
  node degree = centrality index
  "node with most neighbors is most central"

Disadvantage:
- Uses only local knowledge
- Nodes connecting subgraphs are not recognized as important, e.g. the middle node in the 3rd and 4th graph.



Which node is more important/central in network?

Network properties
Centrality

Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
  - Draw the shortest paths for all node pairs: How many of those are going through v_i?

NTNU Norwegian University of Science and Technology

Network properties
Centrality

Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
  - Draw the shortest paths for all node pairs: How many of those are going through v_i?

$$\sum_{s<t \wedge (s,t \neq i)} n^i_{s,t} / n_{s,t}$$

$n^i_{s,t}$: # shortest paths between v_s and v_t going through v_i

$n_{s,t}$: # shortest paths between v_s and v_t

Note: a node pair may have several shortest paths (of equal length)

$n^i_{s,t} / n_{s,t}$

between 0 or 1 (if only one shortest path: either 0 or 1)

NTNU Norwegian University of Science and Technology

Network properties
Centrality

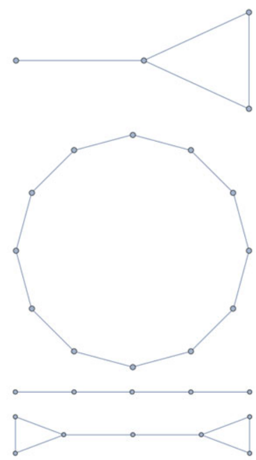
Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
  - Draw the shortest paths for all node pairs: How many of those are going through v_i?

$$\sum_{s<t \wedge (s,t \neq i)} n^i_{s,t} / n_{s,t}$$

$n^i_{s,t}$: # shortest paths between v_s and v_t going through vi

$n_{s,t}$: # shortest paths between v_s and v_t



Network properties
Centrality

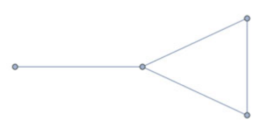
Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
  - Draw the shortest paths for all node pairs: How many of those are going through v_i?

$$\sum_{s<t \wedge (s,t \neq i)} n^i_{s,t} / n_{s,t}$$

$n^i_{s,t}$: # shortest paths between v_s and v_t going through vi

$n_{s,t}$: # shortest paths between v_s and v_t



The left node: 0 - No shortest path of other node pairs goes through it.

Middle node: 2 (How to calculate?)

Right top node: 0

Right bottom node: 0

Network properties
Centrality

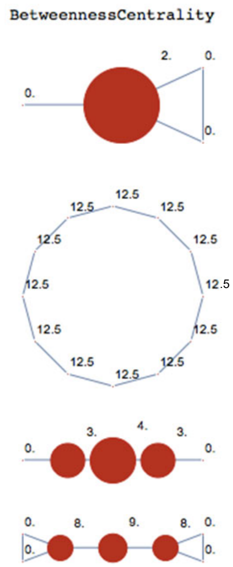
Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
  - Draw the shortest paths for all node pairs: How many of those are going through v_i?

$$\sum_{s<t\wedge(s,t\neq i)} n^i_{s,t}/n_{s,t}$$

$n^i_{s,t}$: # shortest paths between v_s and v_t going through vi

$n_{s,t}$: # shortest paths between v_s and v_t



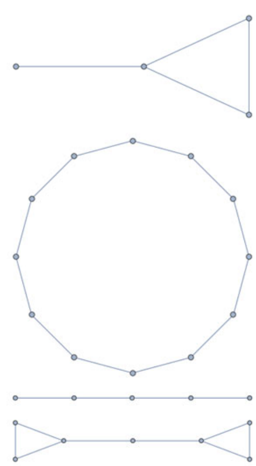
A

NTNU Norwegian University of Science and Technology

Network properties
Centrality

Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
- **Closeness Centrality**
  - How close is a node with all the other nodes?



NTNU Norwegian University of Science and Technology

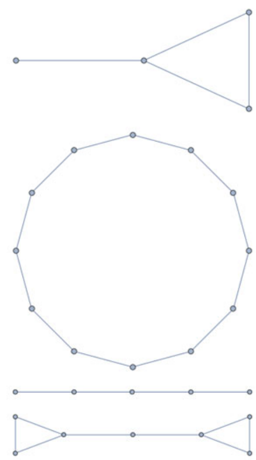
Network properties Centrality

Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
- **Closeness Centrality**
  - How close is a node with all the other nodes?
  - Based on sum of shortest path between v_i and all other nodes

$$\frac{n-1}{\sum_{j \neq i} d_{i,j}}$$

$d_{i,j}$: shortest path distance btw. v_i and v_j
n : # nodes in network



NTNU Norwegian University of Science and Technology

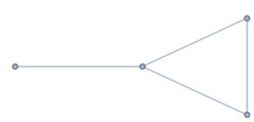
Network properties Centrality

Different centrality measures:

- **Degree Centrality**
- **Betweenness Centrality**
- **Closeness Centrality**
  - How close is a node with all the other nodes?
  - Based on sum of shortest path between v_i and all other nodes

$$\frac{n-1}{\sum_{j \neq i} d_{i,j}}$$

$d_{i,j}$: shortest path distance btw. v_i and v_j
n : # nodes in network



The left node: shortest path distances $d_{i,j}$ *(1, 2, 2) -> 3/5*

Middle node: (1,1,1) -> 3/3

Right top node: (2,1,1) -> 3/4

Right bottom node: (2,1,1) -> 3/4

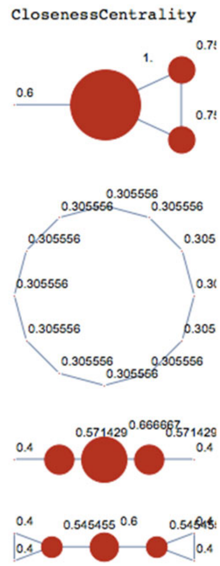
NTNU Norwegian University of Science and Technology

Network properties
Centrality

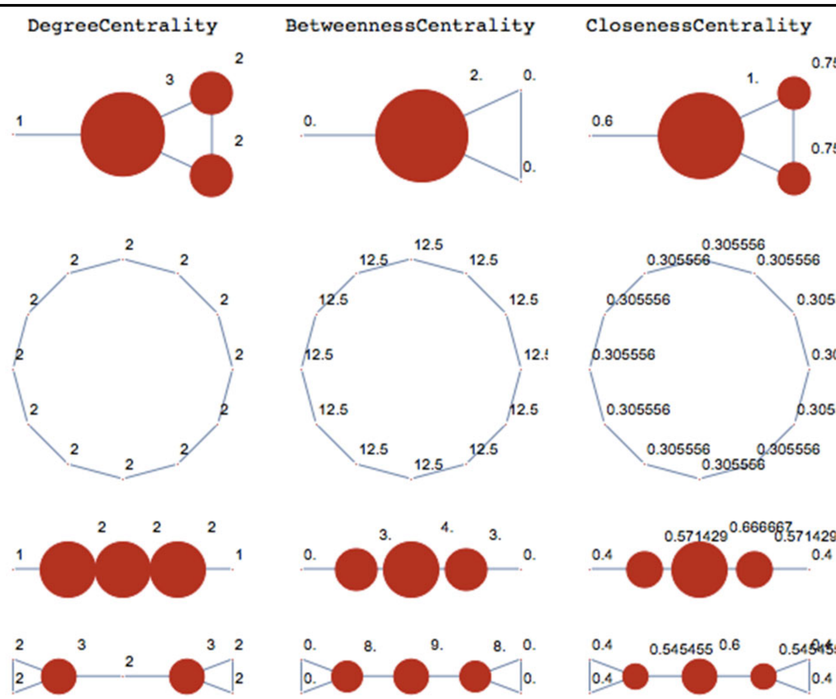
Different centrality measures:
- **Degree Centrality**
- **Betweenness Centrality**
- **Closeness Centrality**
  - How close is a node with all the other nodes?
  - Based on sum of shortest path between v_i and all other nodes

$$\frac{n-1}{\sum_{j \neq i} d_{i,j}}$$

$d_{i,j}$: distance between v_i and v_j
n : # nodes in network



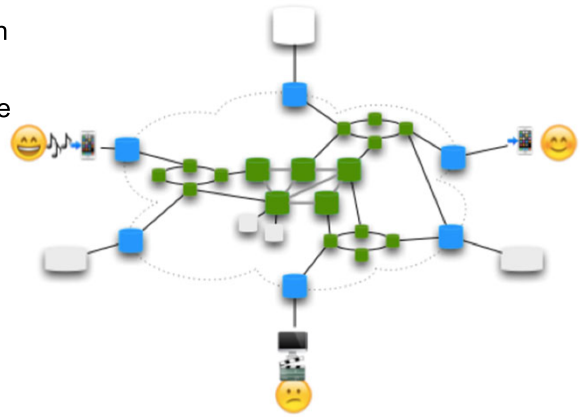
NTNU Norwegian University of Science and Technology



Network Properties
Centrality: Example

Topology for different parts of the network

- Core – mesh
- Metro – ring
- Access - tree



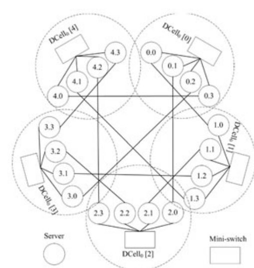
NTNU Norwegian University of Science and Technology

Network Properties
Example

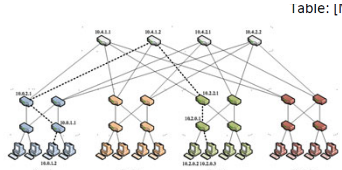
Feature	*DCell*	*FatTree*	*ThreeTier*
$\lvert N \rvert$	2709	2500	2562
$\lvert E \rvert$	4515	6000	2740
$\langle k \rangle \pm \sigma$	3.33 ±0.94	4.8 ±7.6	2.13 ±4.64
$k_{\max}$	4	20	40

k: node degree

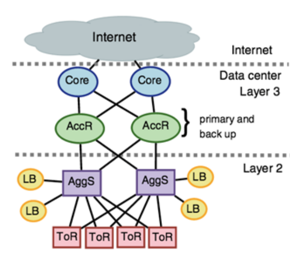
Table: [Manzano 2013]



D-Cell Topology



Fat Tree Topology



Three-Tier Topology

Source: Manzano 2013

NTNU Norwegian University of Science and Technology

Network Properties Centrality: Example

- Social network in *A Storm of Swords,* third book of *A Song of Ice and Fire* (Game of Thrones)

1. Model graph/network
   - Vertices: Characters
   - Edges: connections, i.e. names appeared within 15 words in book
2. Analyze centrality
3. Map of characters, families and their relations purely based on text analysis -> *Find who is the most important character in book.*

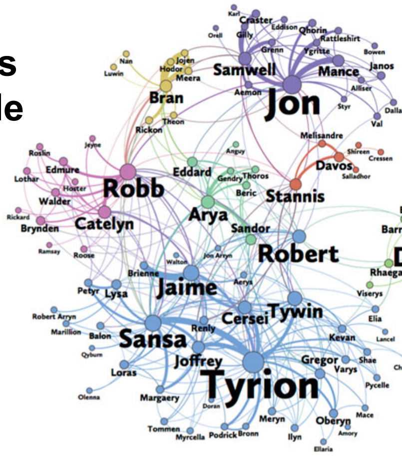


Figure 2. The social network generated from *A Storm of Swords.* The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

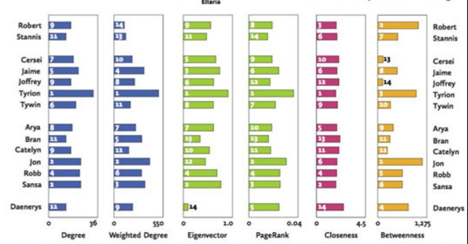
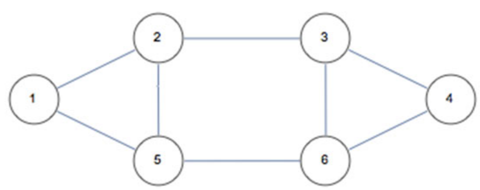


Figure 3. Centrality measures for the network. Larger values correspond to greater importance, except for closeness centrality, where smaller values are better. Numbers in the bars give the rankings of these characters.

Network of Thrones by Andrew Beveridge and Jie Shan, Math Horizons

Flow Analysis (for self-interest)

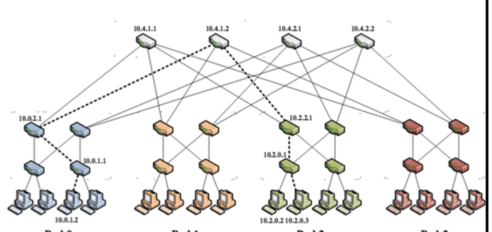
- Each link has a capacity
  - Find maximum flow/capacity between two nodes
- Analyze the needed capacity of links



Assume unit capacity for each link. Then, the maximum flow from node 1 to node 4 is 2: The capacity of the minimum cut, or for any cut, the minimum capacity is 2.

Example application:

- Data Center Network



Max flow problem: The maximum value of an s-t flow is equal to the minimum capacity of an s-t cut (i.e., cut severing s from t) in the network, as stated in the max-flow min-cut theorem.

Image: [Al-Fares 2008]

Agenda

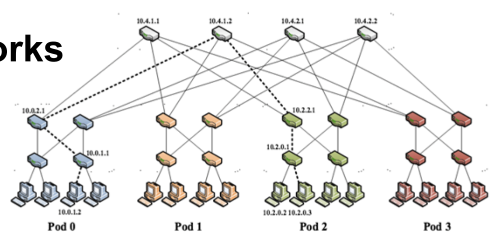
- About Graphs (Repetition)

Structural Analysis of graphs/networks

- Network Properties
  – Basic Network Properties
  – Node Degree (and Random Networks)
  – Centrality
  – Flow Analysis
- Dependability analysis in networks
  – Robustness

Robustness
Dependability in networks

- How to assess dependability in a network?
- **Network properties**
  – Centrality
  – Average node degree
  – …



Feature	FatTree
$\|N\|$	2500
$\|E\|$	6000
d	0.00192
$\langle k\rangle \pm \sigma$	4.8 ± 7.6
λ_1	17.4186
$\mu_{\|N\|-1}$	0.31528
$k_{\max}$	20
κ	1
ρ	1
$\langle l\rangle \pm \sigma$	5.21 ± 1.12
D	6
$\langle b\rangle \pm \sigma$	0.002 ± 0.003
r	-0.2

d: density value
<k>: avg. nodal degree
λ1: spectral radius (largest eigenvalue of the adjacency matrix)
node (κ) and link (ρ) connectivity → single node or link failure may cause network fragmentation
<l>: avg. shortest path length
D: Diameter (longest shortest path)
<b>: avg. node betweenness centrality

Table: [Manzano 2013]

Image: [Al-Fares 2008]

Robustness
Dependability in networks

- How to assess dependability in a network?
- Network properties
- **Robustness**
  - Study impact of node/link removal
  - Remove gradually nodes/links, study how network changes
  - Assess network based on connectivity, centrality, maximum flow, etc.
  - In network science: **often *biggest component* is used (i.e. # nodes in biggest surviving network piece/component)**

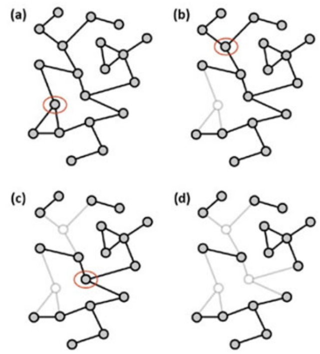
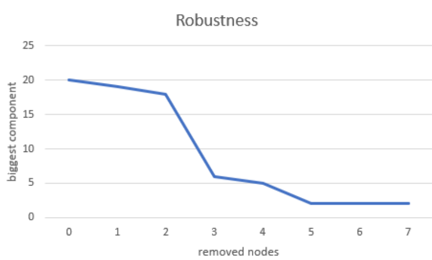
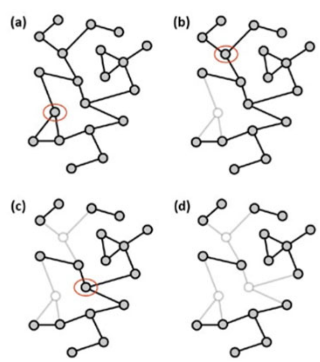


Figure: ***Network Science*** by Albert-László Barabási

NTNU Norwegian University of Science and Technology

Robustness
Dependability in networks





NTNU Norwegian University of Science and Technology

Robustness
Scale-Free networks

- Large scale-free network (generated with Barabási-Albert algorithm)
- Depending on parameters, the change is more abrupt

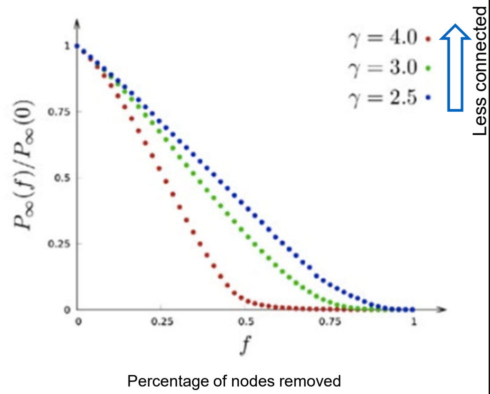


Figure: ***Network Science*** by Albert-László Barabási

Robustness
Scale-Free networks

- Networks react very differently to random failures and targeted attacks
- Here: attack takes out nodes with highest node degree

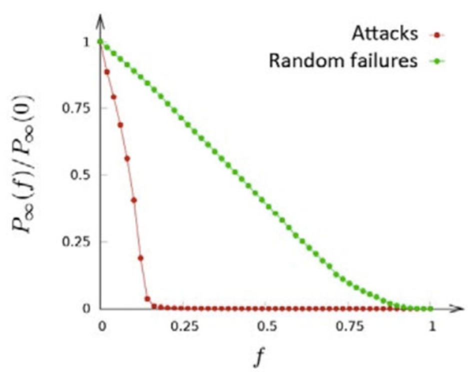
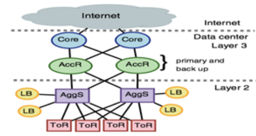


Figure: ***Network Science*** by Albert-László Barabási

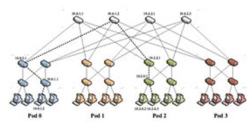
Robustness
Use Case: Data Center Networks

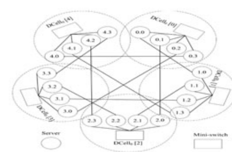
Assume: Each node wants to connect to all other nodes: how many of all the connections are still there after a failure?



Three-Tier Topology

- Robustness with another measure: **A2TR**
  - Two-terminal reliability (2TR): 1 if there exists a path between the two terminals; 0 otherwise.
  - A2TR(p): Average 2TR, i.e., *fraction of node pairs that are connected to each other after p failures* [Neumayer 2010]
  - Fully connected: A2TR=1

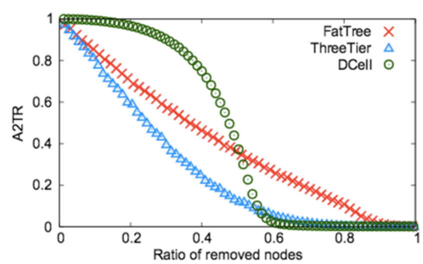

Fat Tree Topology


D-Cell Topology

Figures: Manzano 2013

Robustness
Use Case: Data Center Networks

- Robustness with measure:
  - A2TR(p): fraction of node pairs that are connected to each other after p failures [Neumayer 2010]
  - Fully connected: A2TR=1
- Random networks
- 1000 simulation runs
- Remove stepwise from 0 to (n-2) nodes



(a) Random

Feature	*DCell*	*FatTree*	*ThreeTier*
$\|N\|$	2709	2500	2562
$\|E\|$	4515	6000	2740
d	0.00123	0.00192	0.00083
$\langle k \rangle \pm \sigma$	3.33 ±0.94	4.8 ±7.6	2.13 ±4.64
λ_1	3.56155	17.4186	10.25044
$\mu_{\|N\|-1}$	0.12439	0.31528	0.02308
$k_{\max}$	4	20	40
κ	1	1	1
ρ	1	1	1
$\langle l \rangle \pm \sigma$	8.51 ±1.93	5.21 ±1.12	5.72 ±0.71
D	15	6	6
$\langle b \rangle \pm \sigma$	0.003 ±0.001	0.002 ±0.003	0.002 ±0.014
r	−0.25	−0.2	−0.8961

Table: [Manzano 2013]

Note from the study:

- FatTree: has better metrics (network properties) like average node degree <k>, etc. than Dcell; But FatTree is worse in connectivity analysis

Figures: Manzano 2013

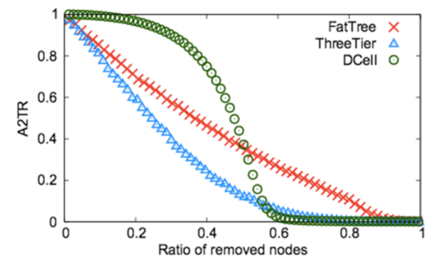
Robustness
Use Case: Data Center Networks

Take-away from the study:

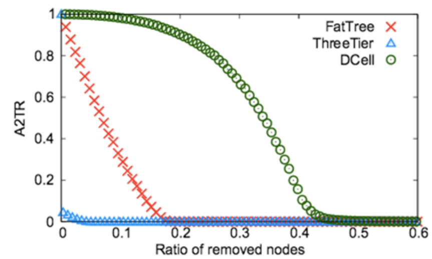
- FatTree: has better metrics (network properties) like average node degree <k>, etc. than DCell;
- But, FatTree is worse than DCell in connectivity analysis

Feature	*DCell*	*FatTree*	*ThreeTier*
$\lvert N\rvert$	2709	2500	2562
$\lvert E\rvert$	4515	6000	2740
d	0.00123	0.00192	0.00083
$\langle k\rangle\pm\sigma$	3.33 $\pm$0.94	4.8 $\pm$7.6	2.13 $\pm$4.64
λ_1	3.56155	17.4186	10.25044
$\mu_{\lvert N\rvert-1}$	0.12439	0.31528	0.02308
$k_{\max}$	4	20	40
κ	1	1	1
ρ	1	1	1
$\langle l\rangle\pm\sigma$	8.51 $\pm$1.93	5.21 $\pm$1.12	5.72 $\pm$0.71
D	15	6	6
$\langle b\rangle\pm\sigma$	0.003 $\pm$0.001	0.002 $\pm$0.003	0.002 $\pm$0.014
r	-0.25	-0.2	-0.8961

Table: [Manzano 2013]



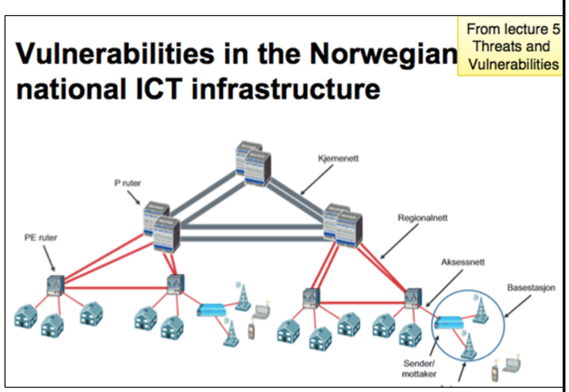
(a) Random



(b) Attack: node degree ee

Robustness
Network Design

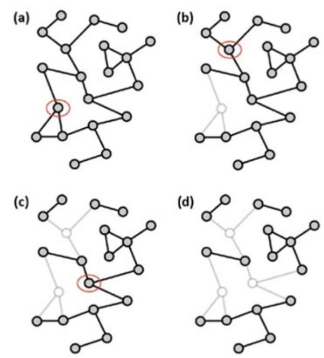
- Network split in several levels
- Each level has different dependability and security requirements
- Structures:
  - Meshed - core
  - Ring - regional
  - Trees - access



Robustness
Summary

- Consider only the structure
  - Rerouting is not considered.
  - Real systems: Coverage not perfect
    - Failures in fail-over mechanism
    - Configuration problems in back up (reroute traffic to failed component)
    - Protocol issues and timeouts

→ Robustness/structural analysis gives the “best" case



Attack: node degree

Summary

- A graph is the abstraction of a real network
- Graph Theory can be used for structural analysis
  - Which nodes are most critical/central (centrality)
  - Find where to invest effort/money
- Methods
  - Network Properties:
    Node degree, centrality
  - *Robustness* (simulate failures of nodes/links)

Literature List

- ***Discrete Mathematics and Its Applications***
  by Kenneth H. Rosen,
  Chapter 10 (pensum in TMA4140 Discrete Mathematics)
- ***Network Science***
  by Albert-László Barabási
  (especially chapter 1-3)
  [online]
  http://barabasilab.neu.edu/networksciencebook/downlPDF.htm
  l

Research on topic:

- [Manzano 2013] **M. Manzano, K. Bilal, E. Calle, and S. U. Khan**, "On the Connectivity of Data Center Networks," IEEE Communications Letters, 2013.
- [Neumayer 2010] **Sebastian Neumayer and Eytan Modiano**, "Network Reliability With Geographically Correlated Failures", Proc. 2010 Conference on Information Communications

About the 3rd exercise

- A tool package based on Python will be used.
- An introduction about the exercise and the tool will be provided.
- Don't forget to use the assigned slots for getting help from assistants (see the course Forelesningsplan.)