# DNA
# Key Metrics

Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

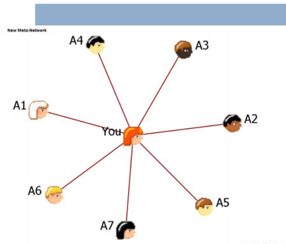
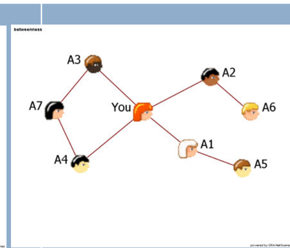
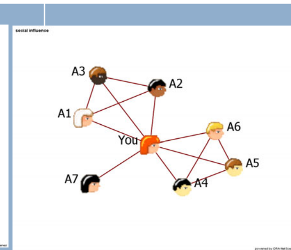
Carnegie Mellon

## Social Informatics

- Who is and will be important?
  - Current Standard – centralities
  - Future – secondary actors, emergent leaders
- What happens if an influential actor is removed?
- What are the core issues ?
  - "talk"
  - Critical words – communicative reach
- What can be done to effect change?
- Who will be the next leader?
- What happens if a group is disbanded?
- Who has influence where?
- How can key actors be influenced?

Carnegie Mellon isr institute for SOFTWARE RESEARCH

## Network Effects

| Access | Control of Flows | Influence |
|---|---|---|
|   |   |   |
| The more people you are connected to the more you can know | The more you are on the path between people the more you can control | The more you are connected to others who are connected to each other the more influence they have on you and you on them |

Carnegie Mellon isr institute for SOFTWARE RESEARCH

## Key Points About Key Actor Metrics

- There is an analogous graph level metric for all node level metrics
- Shortest path metrics have poor scale properties
- Local versus global influence
  - "atrophication of influence"
- Metrics are influenced by size and density
- Metrics may or may not take weighted links into account
- Theoretically - Social capital, homophilly and power underlie key actor metrics

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Overview on Metrics

- Level
  - Node level
  - Dyad level
  - Graph level
- Node level
  - Direct
    - E.g. degree
  - Path based
    - E.g. betweenness
  - Iterative
    - E.g. page rank
- Graph level
  - Cohesive
    - E.g. density
  - Spread
    - E.g. characteristic path length
  - Lumpiness
    - E.g. clustering coefficient
  - Min, max, mean, std. dev of node level metrics
- 2 (and n) mode metrics
  - Folding
  - Meta-networks



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Graph Level Metrics

## Common Graph Level Metrics

- Metric
- Size
- Link count
- Density
- Isolate count
- Component count
- Reciprocity
- Characteristic path length
- Clustering coefficient



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

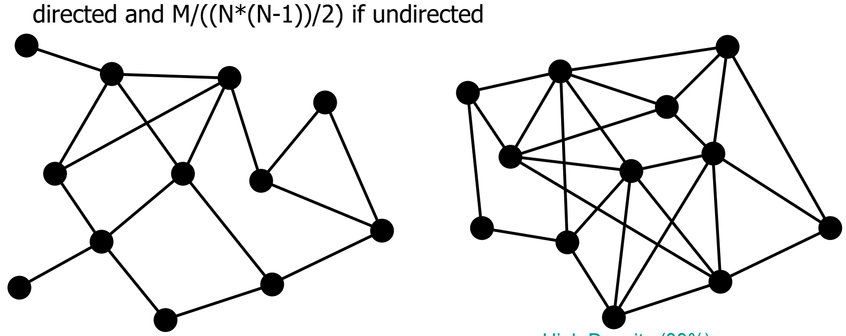
## Size

- Number of nodes (people) in the network
- Matters because as size increases
  - Density decreases
  - Clustering increases
- Reflects network boundary
- Should always be included as a covariate

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Density

- Number of ties, expressed as percentage of the number of ordered/unordered pairs
- Number of ties / Number of possible ties
- If number of nodes = N and number of ties is M, then M/(N*(N-1)) if directed and M/((N*(N-1))/2) if undirected



Low Density (25%)
Avg. Dist. = 2.27

High Density (39%)
Avg. Dist. = 1.76



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Density & Size are Negatively Correlated

- In STEP study we have data from 24 coalitions at baseline
- We correlated size and density and discovered a negative association as predicted:
- R=-0.69

Carnegie Mellon

## Reciprocity (Mutuality, Symmetry)

- Mutual ties: A → B then B→A
- Some relations are inherently symmetric or asymmetric
  - Who did you have lunch with?
  - Who did you go to for advice?
- Reciprocity is calculated as the percent of ties that are reciprocated:

$$R = \frac{(A_{ij} = 1) \& (A_{ji} = 1)}{(A_{ij} = 1) or (A_{ji} = 1)}$$



Carnegie Mellon

## Characteristic Path Length

- Also referred to as average path length
- The average distance from a specific node *i* to all other nodes in the network is defined naturally as

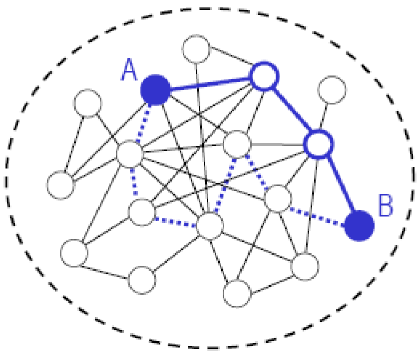
$$\bar{d}(i) = 1/(n-1)\sum_{j=1}^{n} d(i,j)$$

  - Where d(i,j) is the geodesic distance between nodes i and j
- The characteristic path length of the network is defined as the average of these over all nodes in the network, or

$$\bar{d} = 1/n \sum_{i=1}^{n} \bar{d}(i)$$

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Average path length




*The path length between A and B is 3*

- the *path length* between two nodes $A$ and $B$ is the smallest number of edges connecting them:

  $l(A, B) = \min l(A, A_1, \ldots A_n, B)$

- the *average path length* of a network over all pairs of $N$ nodes is

  $L = \langle l(A, B) \rangle$

  $= 2/N(N-1) \sum_{A,B} l(A, B)$

- the *network diameter* is the maximal path length between two nodes:

  $D = \max l(A, B)$

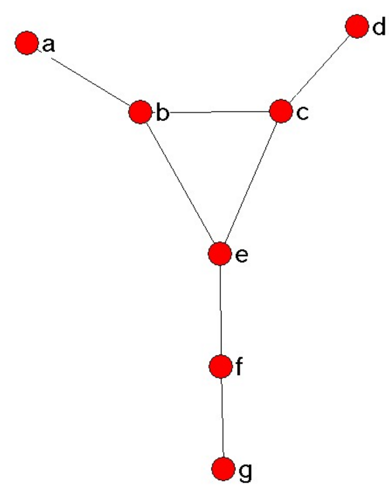
- property: $1 \leq L \leq D \leq N-1$



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Geodesic Distance Matrix

| | a | b | c | d | e | f | g |
|---|---|---|---|---|---|---|---|
| a | 0 | 1 | 2 | 3 | 2 | 3 | 4 |
| b | 1 | 0 | 1 | 2 | 1 | 2 | 3 |
| c | 2 | 1 | 0 | 1 | 1 | 2 | 3 |
| d | 3 | 2 | 1 | 0 | 2 | 3 | 4 |
| e | 2 | 1 | 1 | 2 | 0 | 1 | 2 |
| f | 3 | 2 | 2 | 3 | 1 | 0 | 1 |
| g | 4 | 3 | 3 | 4 | 2 | 1 | 0 |

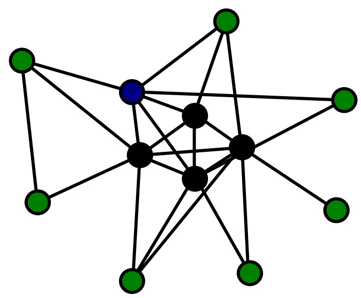


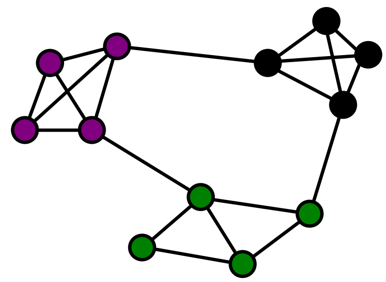

 

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Average Distance

- Average geodesic distance between all pairs of nodes





Core/Periphery
c/p fit = 0.97, avg. dist. = 1.9

Clique structure
c/p fit = 0.33, avg. dist. = 2.4

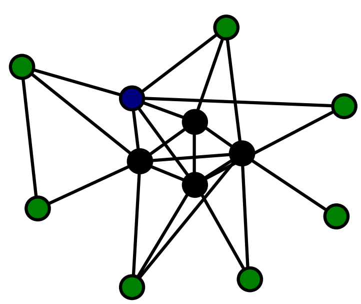


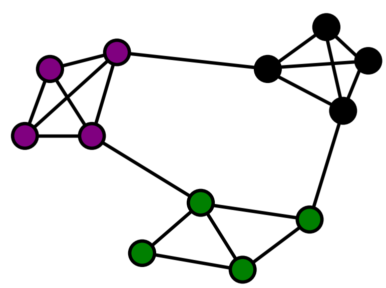


Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Diameter

- Maximum distance between any pair of nodes





Diameter = 3

Diameter = 4

# Clustering Coefficient

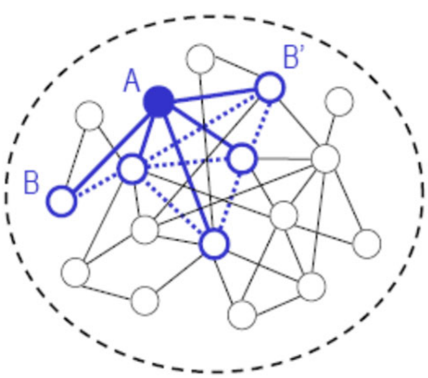
- A measure of degree to which nodes in a graph tend to cluster together
- Defined as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$





# Clustering coefficient




*The clustering coefficient of A is 0.6*

- the *neighborhood* of a node $A$ is the set of $k_A$ nodes at distance 1 from $A$
- given the number of *pairs* of neighbors:

$$F_A = \sum_{B,B'} 1 = k_A(k_A - 1)/2$$

- and the number of pairs of neighbors that are also *connected* to each other:

$$E_A = \sum_{B \leftrightarrow B'} 1$$

- the *clustering coefficient* of $A$ is

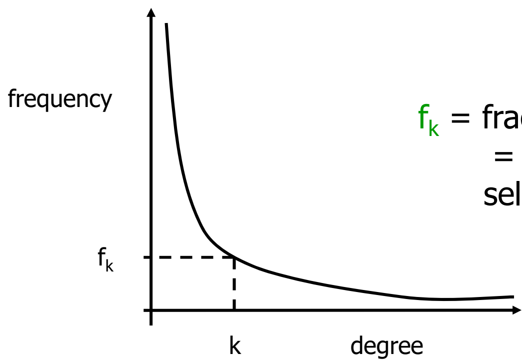
$$C_A = E_A / F_A \le 1$$

- and the *network clustering coefficient*:

$$\langle C \rangle = 1/N \sum_A C_A \le 1$$

Carnegie Mellon
ISr institute for SOFTWARE RESEARCH

# Degree distributions



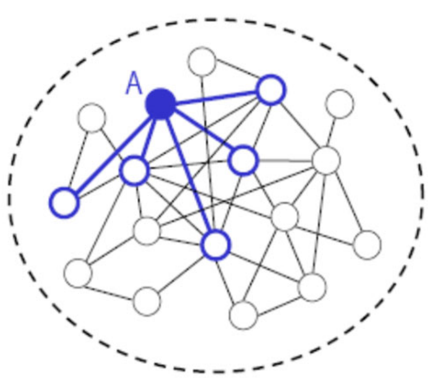

$f_k$ = fraction of nodes with degree $k$
= probability of a randomly selected node to have degree $k$

We are often interested in finding the probability distribution that best fits the observed data



Carnegie Mellon
ISr institute for SOFTWARE RESEARCH

# Degree distribution (connectivity)



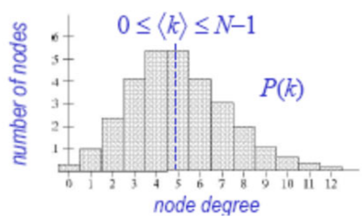

*The degree of A is 5*

- the *degree* of a node $A$ is the number of its connections (or neighbors), $k_A$
- the *average degree* of a network is

$$\langle k \rangle = 1/N \sum_A k_A$$

- the *degree distribution* function $P(k)$ is the histogram (or probability) of the node degrees: it shows their spread around the average value

$$0 \leq \langle k \rangle \leq N\text{–}1$$

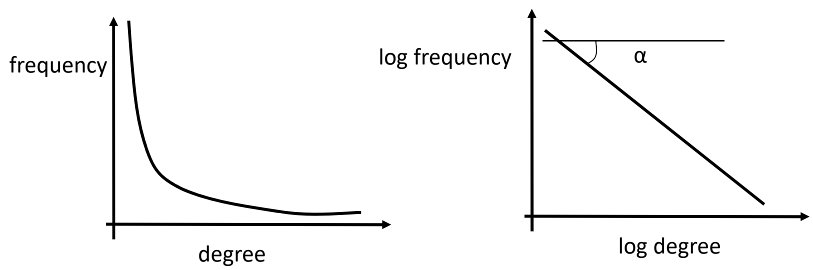


Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Power-law of Total Degree

- Power-law distribution gives a line in the log-log plot

$$\log p(k) = -\alpha \log k + \log C$$




- α : power-law exponent (typically $2 \le \alpha \le 3$)



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Other Key Graph Level Metrics

- Average Degree Centrality
- Average Betweenness
- Average EigenVector Centrality
- …
- Standard Deviation of …
- Assortativity
- Modularity
- Factions
- …

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Node Level Metrics



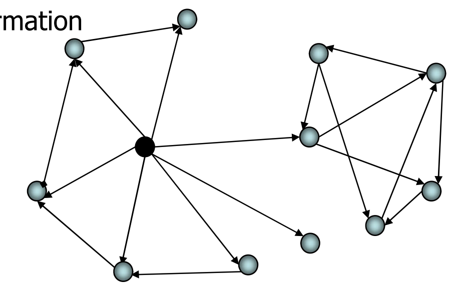
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## Power & Position

- Many network metrics are designed to capture power or influence
  - Betweenness, Degree, Information-Centrality …
- Good – but don't go far enough
  - These measures are frequently highly correlated
    - Often due to large number of individuals who are equivalently low
  - Individuals who are top are often obvious
    - e.g. the president or CEO
    - Particularly true with political elite data gathered from news
- Extend ability to identify powerful individuals
  - E.g., the emergent leader
  - E.g., the power behind the throne

# Network Elite

- Nodes that stand out as high/low on some measure
- Power
  - Bonacich power centrality = out-degree (row) centrality when beta = 0
  - Access to resources, information, people
  - Ability to mobilize others (reach)
  - Ability to control the flow of information
  - Ability to give orders
  - Ability to broker between groups





# Identifying Network Elite

- Centrality Approach
  - How much matters
- Brokerage
  - Who you connect matters
- But …
  - It matters what is flowing through the network
- It matters if network is multi-mode, multi-plex, multi-way

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Centralities

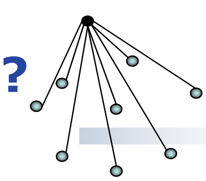
- Degree Centrality
  - Node with the most connections
- Betweenness Centrality
  - Node in the most best paths
    - Requires symmetric data in some tools but not ORA
- Eigenvector Centrality
  - Node connected best overall
    - Doesn't work if there are components in some tools
- Closeness Centrality
  - Node that is closest to all other nodes

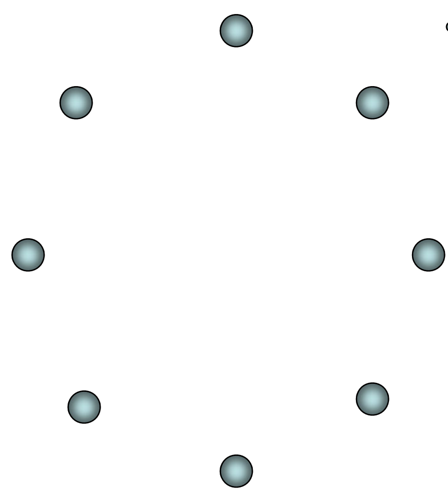
*Issue: Measures are highly correlated*



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Who Is "Key" ?



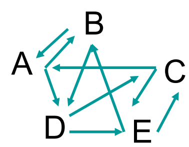


Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Degree Centrality

- Degree – total number of edges/ nodes ego is connected to
  - Commonly thought of as a measure of influence or importance
- In Degree – total number of nodes that send edge to ego (column)
- Out Degree – total number of nodes that receive edge from ego (row)
- Sink – 0 in degree; Source – 0 out degree

$$\begin{matrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{matrix}$$



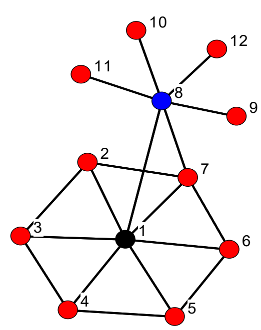

| N | In | Out | Total |
|---|---|---|---|
| A | 2 | 2 | 4 |
| B | 2 | 2 | 4 |
| C | 2 | 2 | 4 |
| D | 2 | 2 | 4 |
| E | 2 | 2 | 4 |



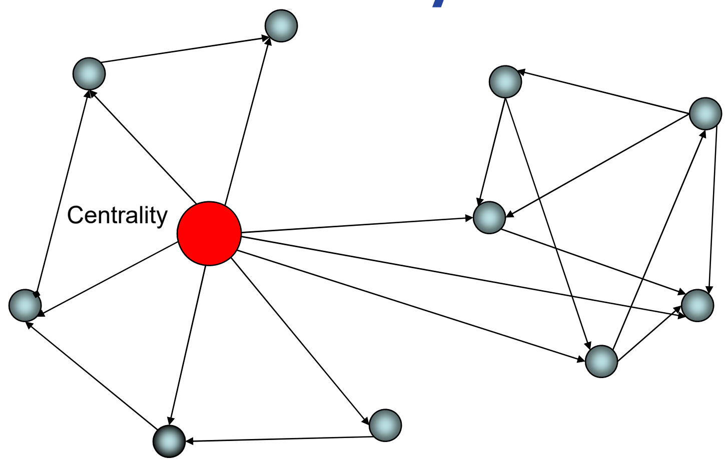
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Degree Centrality

- Number of edges incident upon a vertex
  - $d_8 = 6$, while $d_{10} = 1$
- Sum of degrees of all nodes is twice the number of edges in graph
- Average degree = density times (n-1)
- Index of exposure to what is flowing through the network
  - Gossip network: central actor more likely to hear a given bit of gossip
- Interpreted as opportunity to influence & be influenced directly
- Predicts variety of outcomes from virus resistance to power & leadership to job satisfaction to knowledge



# Who Is "Key" ?








# Betweenness Centrality

- Frequency with which a node lies along the shortest path between two other nodes
- Computed as: $$b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}}$$

  where $g_{ij}$ is number of geodesic paths from i to j and $g_{ikj}$ is number of those paths that pass through k
- Index of potential for gate-keeping, brokering, controlling the flow, and also of liaising otherwise separate parts of the network
- Interpreted as indicating power and access to diversity of what flows; potential for synthesizing
- Sometimes interpreted as "connecting" groups
- Very "expensive" to compute

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## Closeness Centrality

- Measured as:
  - Sum of distances to all other nodes
  - Computed as marginals of symmetric geodesic distance matrix
- Closeness is an inverse measure of centrality
- Index of expected time until arrival for given node of whatever is flowing through the network
  - Gossip network: central player hears things first



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## Eigenvector Centrality

- A node will have a high score if it is connected to many nodes that are themselves highly connected
- Computed as:

$$\lambda v = Av$$

where A is adjacency network and V is eigenvector centrality. V is the principal eigenvector of A

- Indicator of popularity and group-bonding
- Like degree, this is an index of exposure, risk
- Tends to identify centers of large cliques
- Often identified as leader of self-contained group, sometimes referred to as leader of leaders
- Very "expensive" to compute

Adapted from Steve Borgatti 2004

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

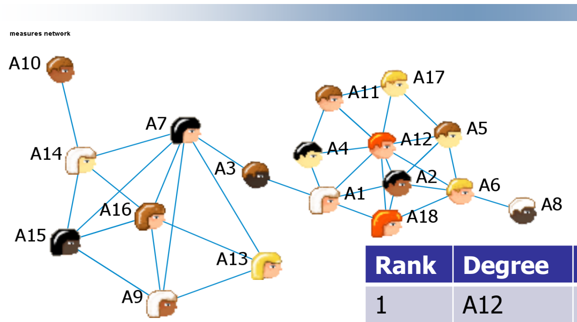
# Characteristics of networks relation to Measures

| Measure | Allow/Ignore Self-Loops | Symmetric/ Asymmetric | Binary/ Weighted | Connected/ Disconnected |
|---|---|---|---|---|
| Degree | yes | yes | yes | no |
| Betweenness | no | yes | yes | no |
| Closeness | no | yes | yes | yes |
| Eigenvector | yes | no | yes | yes |
| Clustering Coefficient | yes | yes | no | no |



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

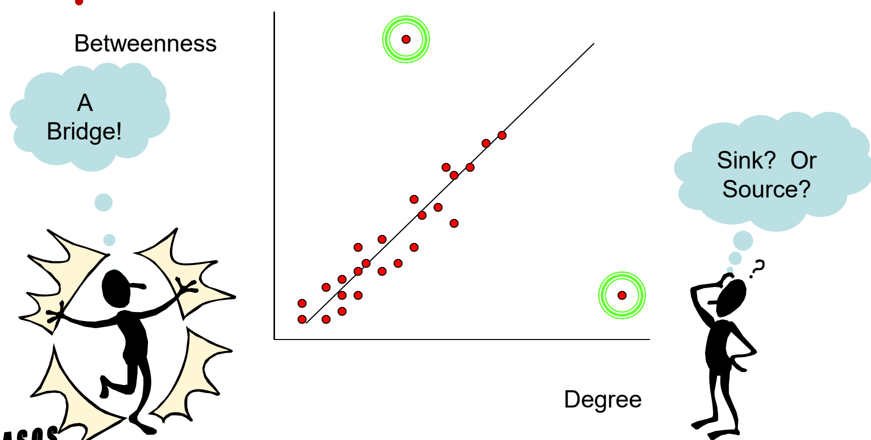
# Illustrative Network




| Rank | Degree | Betweenness | Eigenvector |
|---|---|---|---|
| 1 | A12 | A1 | A12 |
| 2 | A7 | A3 | A2 |
| 3 | A1, A2, A16, A6 | A7 | A6 |
| 4 | A14, A15, A18, A3, A5 | A12 | A18 |

# Moving Beyond Single Measures

***Issue: Centrality Measures are highly correlated***





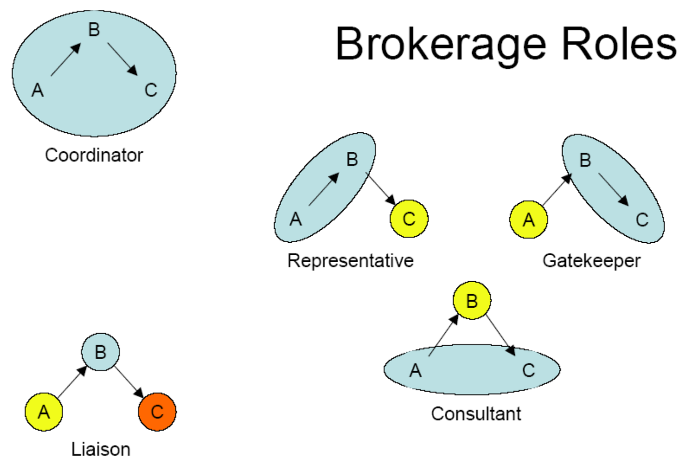

# Simple SNA Measures

| Measure | Definition | Meaning | Usage |
|---|---|---|---|
| Degree Centrality | Node with the most connections | In the know | Identifying sources for intel; Reducing information flow |
| Betweenness | Node in the most best paths<br>Needs symmetric data | Connects groups | Typically has political influence, but may be too constrained to act |
| Eigenvector centrality | Node most connected to other highly connected nodes | Strong social capital | Identifying those who can mobilize others |
| Closeness | Node that is closest to all other nodes | Rapid access to all information | Identifying sources to acquire/transmit information |
| Betweenness - Centrality | High in betweenness but not degree centrality | Connects disconnected groups | Go-between; Reduction in activity by disconnecting groups |

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Brokerage Connections Among …

Brokerage Roles



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Brokerage Common ways to measure

Common ways to Measure

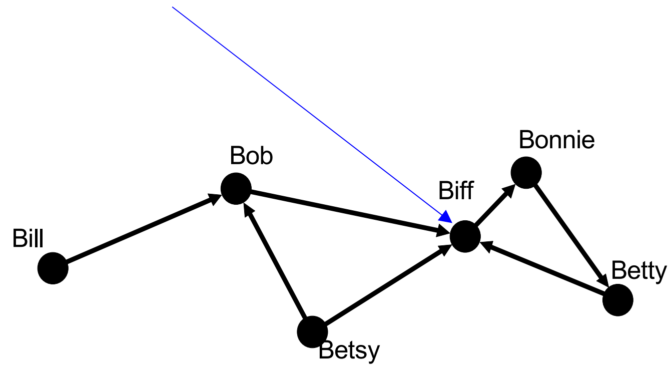
- Cutpoints
- Bridges
- Structural holes
- Embeddedness in triads
- Embeddedness in cliques

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Cutpoints

- Nodes which, if deleted, would disconnect net

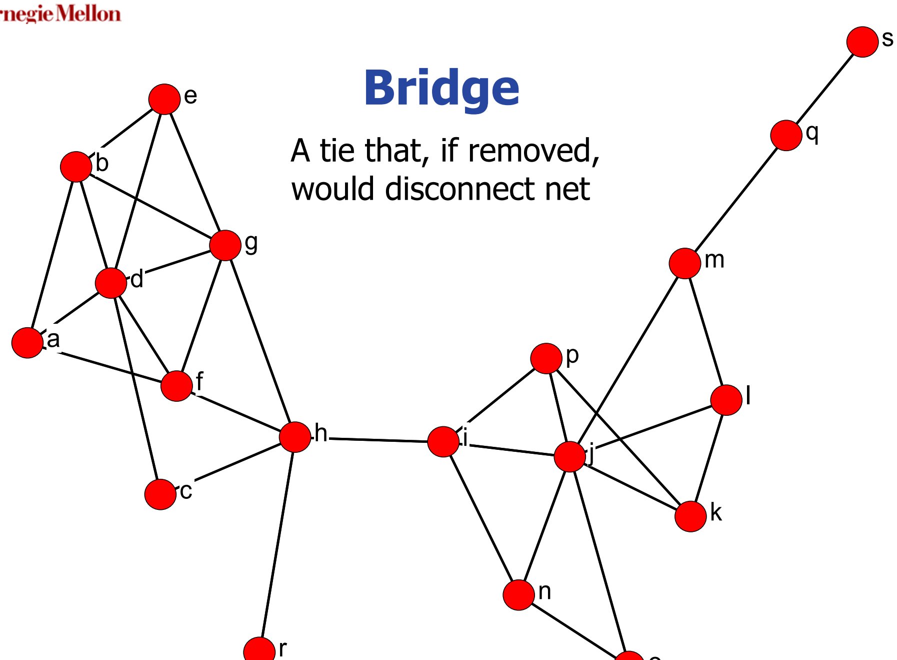


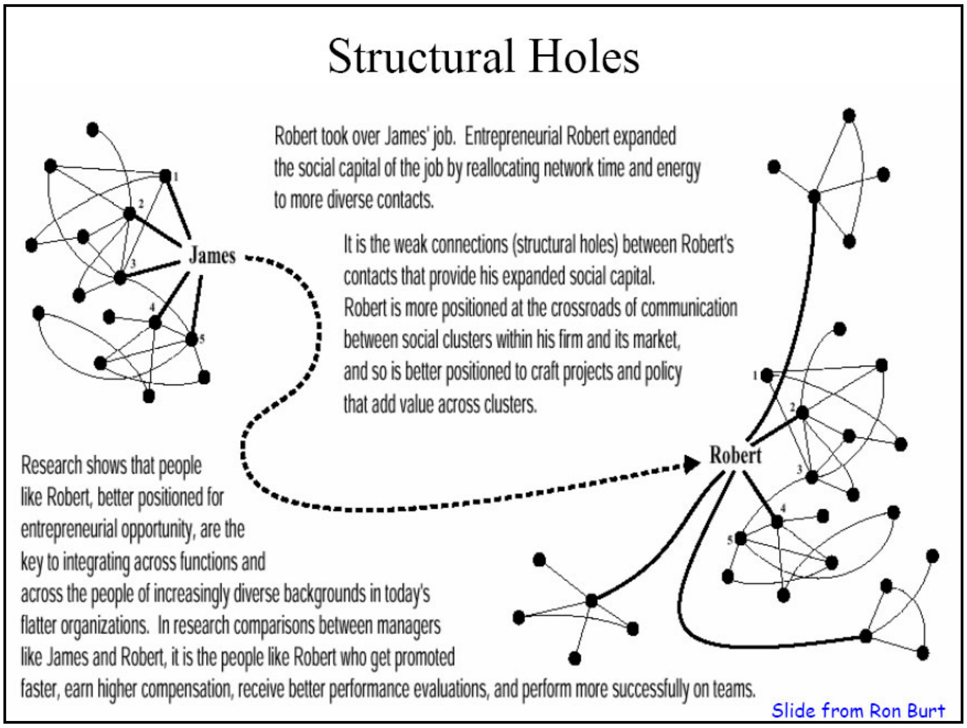
Carnegie Mellon
isr

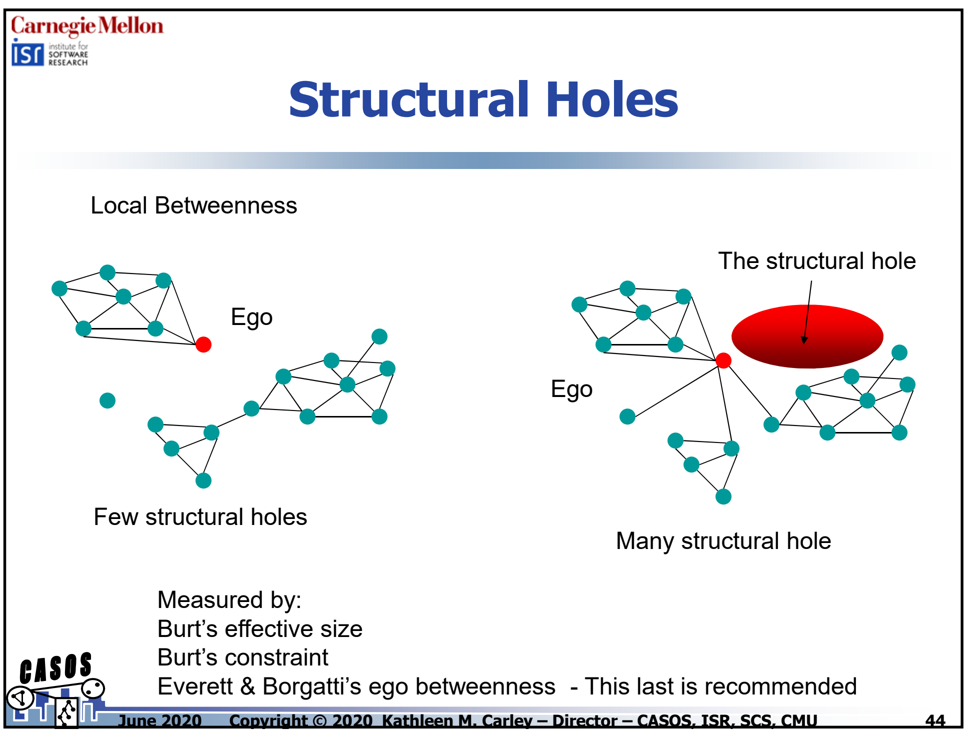
# Bridge

A tie that, if removed, would disconnect net



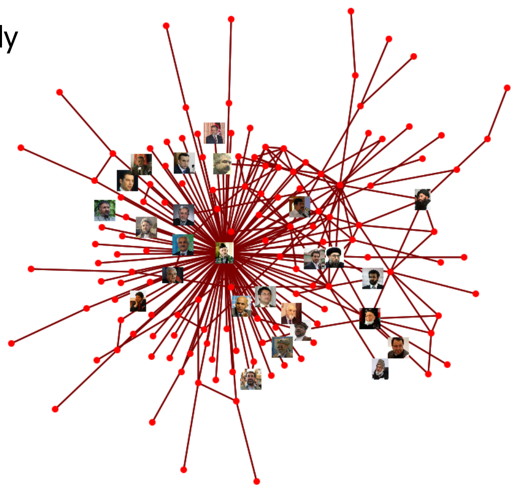

Structural Holes
Robert took over James' job. Entrepreneurial Robert expanded the social capital of the job by reallocating network time and energy to more diverse contacts.
It is the weak connections (structural holes) between Robert's contacts that provide his expanded social capital.
Robert is more positioned at the crossroads of communication between social clusters within his firm and its market, and so is better positioned to craft projects and policy that add value across clusters.
James
Robert
Research shows that people like Robert, better positioned for entrepreneurial opportunity, are the key to integrating across functions and across the people of increasingly diverse backgrounds in today's flatter organizations. In research comparisons between managers like James and Robert, it is the people like Robert who get promoted faster, earn higher compensation, receive better performance evaluations, and perform more successfully on teams.
Slide from Ron Burt


Carnegie Mellon
isr institute for SOFTWARE RESEARCH
Structural Holes
Local Betweenness
Ego
Few structural holes
The structural hole
Ego
Many structural hole
Measured by:
Burt's effective size
Burt's constraint
Everett & Borgatti's ego betweenness - This last is recommended
CASOS
June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU
44

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Familiar Secondary Actors

- Leaders
- Latent leaders – most likely to sway populace when leaders removed
- Gatekeepers
  - Betweenness
  - Even better – high betweenness low degree
  - Individuals with high structural holes
- Critical for impacting
  - Who has access to what information
  - Who gets what job
  - Etc.





Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Secondary Actors – One Mode Data

- Gatekeepers
  - Access to information, jobs,
  - High betweenness and low degree centrality
- Power Behind the Throne
  - Ability to mobilize
  - Node – structurally most similar to ego
    - Absolute and relative similarity
    - Number of "contacts" in common
- Latent Leader
  - Strong if current leader is removed
  - High degree after ego is removed based on degree centrality

Carnegie Mellon

# Two Mode Metrics

- Many types of 2 Mode Metrics
  - Quantity
  - Variance
  - Correlation
  - Specialization
- Many type of N Mode Metrics
  - Quantity
  - Coherence
  - Substitution
  - Control



Carnegie Mellon

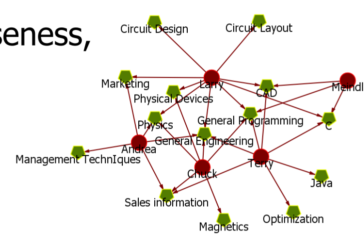
# Two-Mode Data

Plus

- Often easier to collect (e.g., co-publishing)
- Two-mode data seems to provide more privacy
- Allow non-human analysis

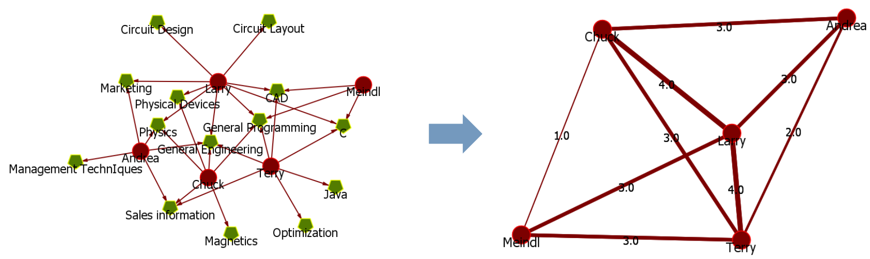
Minus

- SNA metrics (betweenness, closeness, eigenvector, etc.) imply “flows”
- Are they network data at all?



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Approach: Folding Networks




| Person | Avg. Shared Knowledge |
|---|---|
| Larry | 3.50 |
| Terry | 3.00 |
| Chuck | 2.75 |
| Andrea | 2.00 |
| Meindl | 1.75 |

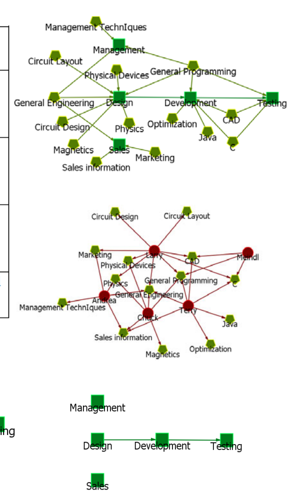
*Faust (1997), Borgatti and Everett (1997)*



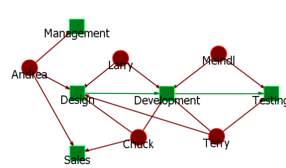
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Approach: Meta-Networks

| Meta-Matrix entities | People | Knowledge/ Resources | Events/ Tasks | Groups/ Organizations |
|---|---|---|---|---|
| People | Social network | Knowledge Network/ Resource Network | Attendance Network/ Assignment Network | Membership network |
| Knowledge/Resources | | Information Network/ Substitution Network | Needs network | Organizational capability |
| Events/Tasks | | | Temporal Ordering/ Task Flow/ Precedence | Institutional support or attack |
| Organizations | | | | Interorganizational network |

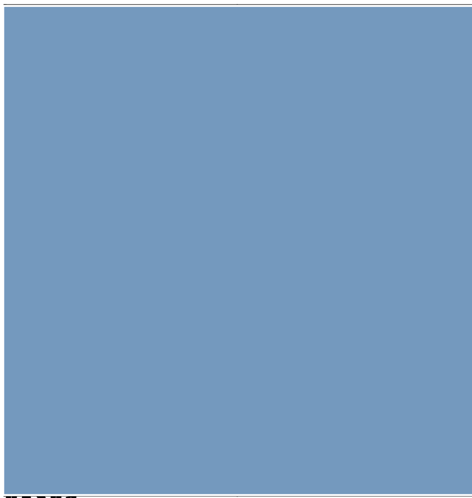


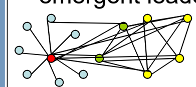
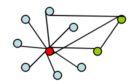
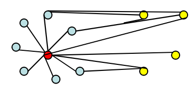
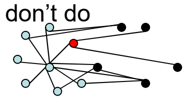

*Krackhardt & Carley (1998)*
*Carley (2002)*



Carnegie Mellon

# Meta-Network KEY ACTORS



| Cognitive Demand | Exclusivity |
|---|---|
| emergent leader  | critical ability  |
| **Redundancy** | **Inverse similarity** |
| backup  | alike in what we don't do  |



Carnegie Mellon

# Many Specific Two-Mode Metrics

- Current ORA measures:
  columnDegreeCentrality, inDegreeCentrality, outDegreeCentrality, rowDegreeCentrality, columnCount, rowCount, edgeCount, capability, knowledgeLoad, resourceLoad, density, rowBreadth, columnBreadth., columnDegreeCentralization, inDegreeCentralization, outDegreeCentralization, rowDegreeCentralization, knowledgeDiversity, resourceDiversity, relativeCognitiveSimilarity, cognitiveSimilarity, relativeSimilarity, correlationSimilarity, relativeCognitiveDistinctiveness, cognitiveDistinctiveness, correlationDistinctiveness, relativeCognitiveResemblance, cognitiveResemblance, correlationResemblance, relativeCognitiveExpertise, cognitiveExpertise, relativeExpertise, correlationExpertise, knowledgeExclusivity, resourceExclusivity, taskExclusivity, exclusivityComplete, exclusivity, columnRedundancy, rowRedundancy, knowledgeRedundancy, accessRedundancy, resourceRedundancy, assignmentRedundancy, knowledgeAccessIndex, resourceAccessIndex,
- Classification: Four concept groups of measures
- Node level + dyad level + network level metrics
- [Knowledge] for any kind of affiliation (events, …)

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# 1. Quantity

Count or average the entries of a matrix

- Degree
  - Node: Counting the row or column entries of a two mode network
  - *Wasserman and Faust (1994), Borgatti and Everett (1997)*

$$\mathbf{d}_i = \frac{\sum_{j=1}^{|K|} AK(i,j)}{|K|}$$

- Load
  - Network: Density, average amount of [knowledge]

$$\mathbf{l} = \frac{\sum AK}{|A|}$$



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Illustration of Quantity: Actual Workload

- The knowledge an agent uses to perform the tasks to which it is assigned.
- Actual Workload for agent 'i' is defined as follows:
  - (AK * KT * AT')(i,i) / sum(KT)
- Input: AK : binary – variable; AT : binary - variable; KT : binary – permanent;
- Output $\Re \in [0,1]$
- Standard Deviation of Actual Workload (AW):

$$S = \sqrt{\sum_{i=1}^{N} (AW_i - \overline{AW})^2 / (N-1)}$$

## 2. Variance

Distribution of connections in networks

- Centralization
  - Network: skewness of node level values
  - *Freeman (1979)*

$$\mathbf{c} = \frac{\sum_{i=1}^{n}[c(p*) - c(p_i)]}{\max \sum_{i=1}^{n}[c(p*) - c(p_i)]}$$

- Diversity
  - Network: Is [knowledge] rather equally distributed or concentrated?
  - *Hirschman (1945), etc.*

$$w_k = \sum_{i=1}^{|A|} AK(i,k)$$

$$W = \sum_{k=1}^{|K|} w_k$$

$$\mathbf{d} = 1 - \sum_{k=1}^{|K|} \left(\frac{w_k}{W}\right)^2$$



## 3. Correlation

Matrix that describe similarities/dissimilarities between all pairs of agents

- Similarity
  - Node: To what degree do other agents have the same [knowledge]?
- Distinctiveness
  - Dyad: Complementary [knowledge]
- Resemblance
  - Dyad: Agents have the exact same knowledge
- Expertise
  - Node: Degree of dissimilarity between agents

$$\mathbf{M} = AK \cdot AK'$$

$$w(i) = \sum \mathbf{M}(i,:) \; for \; 1 \le i \le |A|$$

$$\mathbf{S} = \mathbf{M}(i,j)/w(i)$$

$$\mathbf{s}_i = \frac{\sum_{j=1, j\neq i}^{|A|} \mathbf{S}(i,j)}{|A| - 1}$$

*Carley (2002)*

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# 4. Specialization

<u>Identify agents that have either exclusive or redundant connections</u>

- Exclusivity
  - Node: Exclusively connected to [knowledge]
  - *Ashworth and Carley (2006)*

$$\mathbf{x}_i = \sum_{j=1}^{|K|} [AK(i,j) \cdot \exp(1 - \sum AK(:,j))]$$

- Redundancy
  - Network: different agents sharing the same knowledge.
  - Carley (2002)
- Access
  - Node: Critical connections to [knowledge]
  - Ashworth and Carley (2006)



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Illustration of Specialization: Redundancy

- Modes: Agents x Tasks
- Average number of redundant agents assigned to tasks. An agent is redundant if there is already an agent assigned to the task.
- Redundancy occurs only when more than one agent is assigned to a task. Define the assignment redundancy for task j as follows:, $d_j = \max\{0, sum(AT(:,j)) - 1\} \; 1 \le j \le |T|$
- Then Assignment Redundancy = $\left(\sum_{j=1}^{|T|} d_j\right) / |T|$

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Illustration of Specialization: Exclusivity Index

- Detects agents who have singular knowledge.
- The Knowledge Exclusivity Index (KEI) for agent i is defined as follows:

$$\sum_{j=1}^{|K|} AK(i,j) * e^{(1-sum(AK(:,j)))}$$

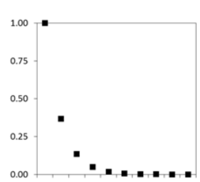
- The values are then normalized to be in [0,1] by dividing by the maximum KEI value.



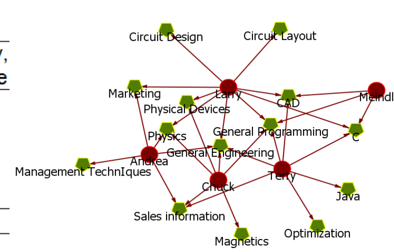
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Exclusivity

- Connection of people with [knowledge] which is shared by no other or at least a small number of other people
- People with high [knowledge] exclusivity are critical
- people with low [knowledge] exclusivity are substitutable
- Company: Knowledge redundancy is 0.286




| Person | Exclusivity, Knowledge |
|---|---|
| Andrea | 0.121 |
| Chuck | 0.124 |
| Larry | 0.232* |
| Meindl | 0.023 |
| Terry | 0.179 |
| AVG | 0.136 |






CASOS

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## Additional Specialized Measures Exist Particularly Ones Using N Mode metrics

- Performance
  - Diffusion
  - Accuracy
- Loads
  - Cognitive demand
  - Workload
  - Potential Work Load
- Congruency – fit
  - Communication
  - Knowledge
  - Resource
- Need for Negotiation
- Under Supply



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## Illustration of Control: Cognitive Demand

- *The cognitive effort the individual has to do on average*
- How many people do you interact with CENTRALITY
- How many tasks do you do
- How much knowledge do you have
- How much knowledge is needed to do the tasks
- How many people do you need to interact with to do the tasks
- How many other tasks and so people depend on you
- How many other tasks and so people do you depend on

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## A Simple Version of PageRank

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$$

- u: a node
- v: a node
- $B_u$: the set of u's in-degree links (v are nodes pointing to u)
- $N_v$: the number of outdegree links of node v
- c: the normalization factor to make $||R||_{L1} = 1$ ($||R||_{L1} = |R_1 + \dots + R_n|$)

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

## K-Shell Decomposition

- Given an undirected graph G=(V,E), k-shell decomposition works in a series of steps iteratively:
- k=1: we start by removing all nodes with degree 1 and the associated edges; assign these nodes to 1-shell
- k=2: we remove all nodes of (remaining) degrees of 2 or less, and the associated edges; assign these nodes to 2-shell
- ….
- k-shell: we remove all nodes of (remaining) degrees of k or less, and the associated edges; these nodes are k-shell nodes
- …
- The process stops when no nodes are left. The last k is kmax

- k-core: the graph formed by the nodes that have not been removed at step k
- k-crust: the graph formed by all the nodes in k'-shells, k'=1, …,k

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# K-Shell Decomposition: k-Shell Index & Network Structures

- Besides its degree, now each node is also assigned a k-shell index
  - denote by shell(v) for a node v; let deg(v) denotes v's degree
  - give us another "bivariate" (or "multivariate") distribution <deg(v),shell(v)>
- Some simple facts:
  - shell(v) <= deg(v) for all v; and clearly if deg(v)=1, shell(v)=1
  - a high degree node may have low k-shell index: for any v w/ arbitrary deg(v)>1, its k-shell index can be as low as 2
  - for v, if the largest degree of its neighbor is d, then shell(v)<=d+1
  - If v is part of s-clique (and thus deg(v)>=s), then shell(v) >= s.
- Connected components in 1-crust: singleton nodes and isolated edges
- Connected components in 2-crust: stars and stars connected via a path
- ……



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

# Summary on Metrics

- Level
  - Node level
  - Dyad level
  - Graph level
- Node level
  - Direct
    - E.g. degree
  - Path based
    - E.g. betweenness
  - Iterative
    - E.g. page rank
- Graph level
  - Cohesive
    - E.g. density
  - Spread
    - E.g. characteristic path length
  - Lumpiness
    - E.g. clustering coefficient
  - Min, max, mean, std. dev of node level metrics
- 2 (and n) mode metrics
  - Folding
  - Meta-networks