Introduction to Complex Networks
and their
Importance to NAADSM

Day #5, Session #5,
NAADSM Course, Fort Collins, Aug'07

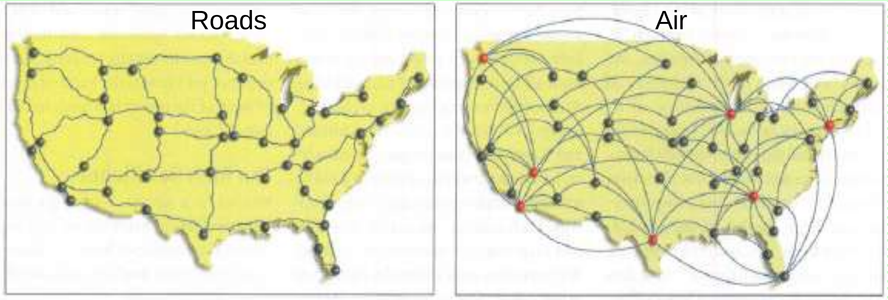
Bruce McNab DVM PhD
Office of the Chief Veterinarian for Ontario, Canada

Networks: Session Outline

- Some examples of networks
- Definitions & tools to depict, describe & analyze networks
- Random vs. scale-free networks
- Implications for spread & control
- Livestock examples of scale-free networks
- NADDSM modeling & networks (now and future)

Examples of Networks

- world wide web
- actors and movies, or social networks
- researchers (even people interested in disease models)
- communications and power systems
  - internet & phone system
  - electrical grid (generators, high-tension, city, within-house)
- transportation systems
  - roads
  - air

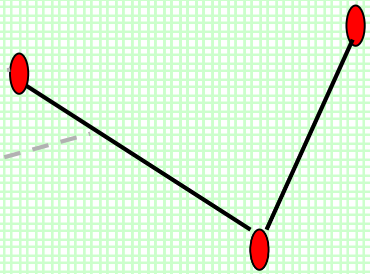


Barabasi & Bonabeau Sci Am '03

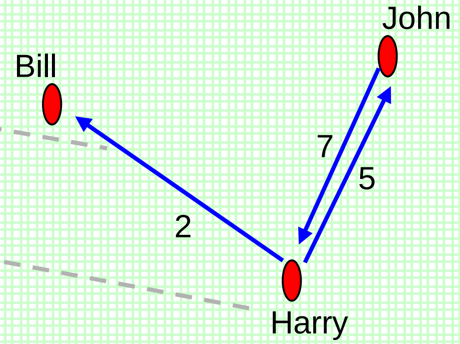
- **farms** (direct and indirect animal contacts)

Definitions, Concepts & Tools

- **graph**
  - **set of vertices & lines between them**



- **node or vertex**
- **lines**
  - **undirected edges**
  - **directed arcs**
- **network**
  - **graph with attribute data**



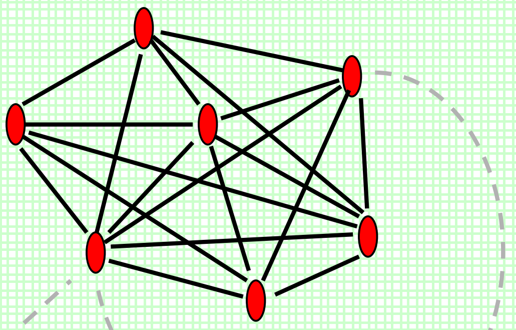
Exploratory Social Network Analyses with Pajek

de Nooy, Mrvar, & Batagelj

Cambridge University Press 2005

Definitions, Concepts & Tools cont.

Highly Connected (undirected)



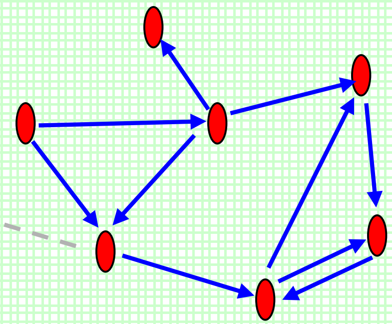
Density of Connectivity

- proportion of all possible connections

Vertex Degree

- number of incident lines
- in-degree & out-degree

e.g. Less Connected (directed)

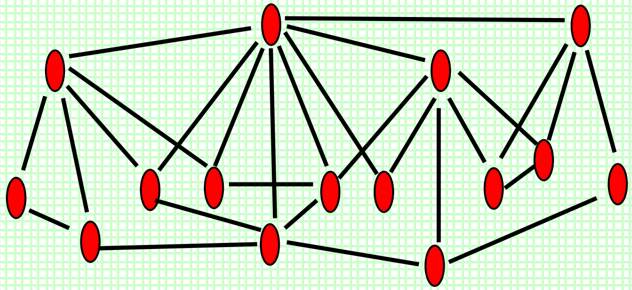


Definitions, Concspets & Tools cont.

One-mode network

• each vertex can relate to any other

club members, farms

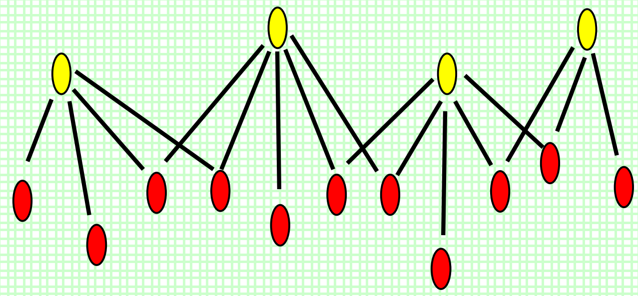


Two-mode network

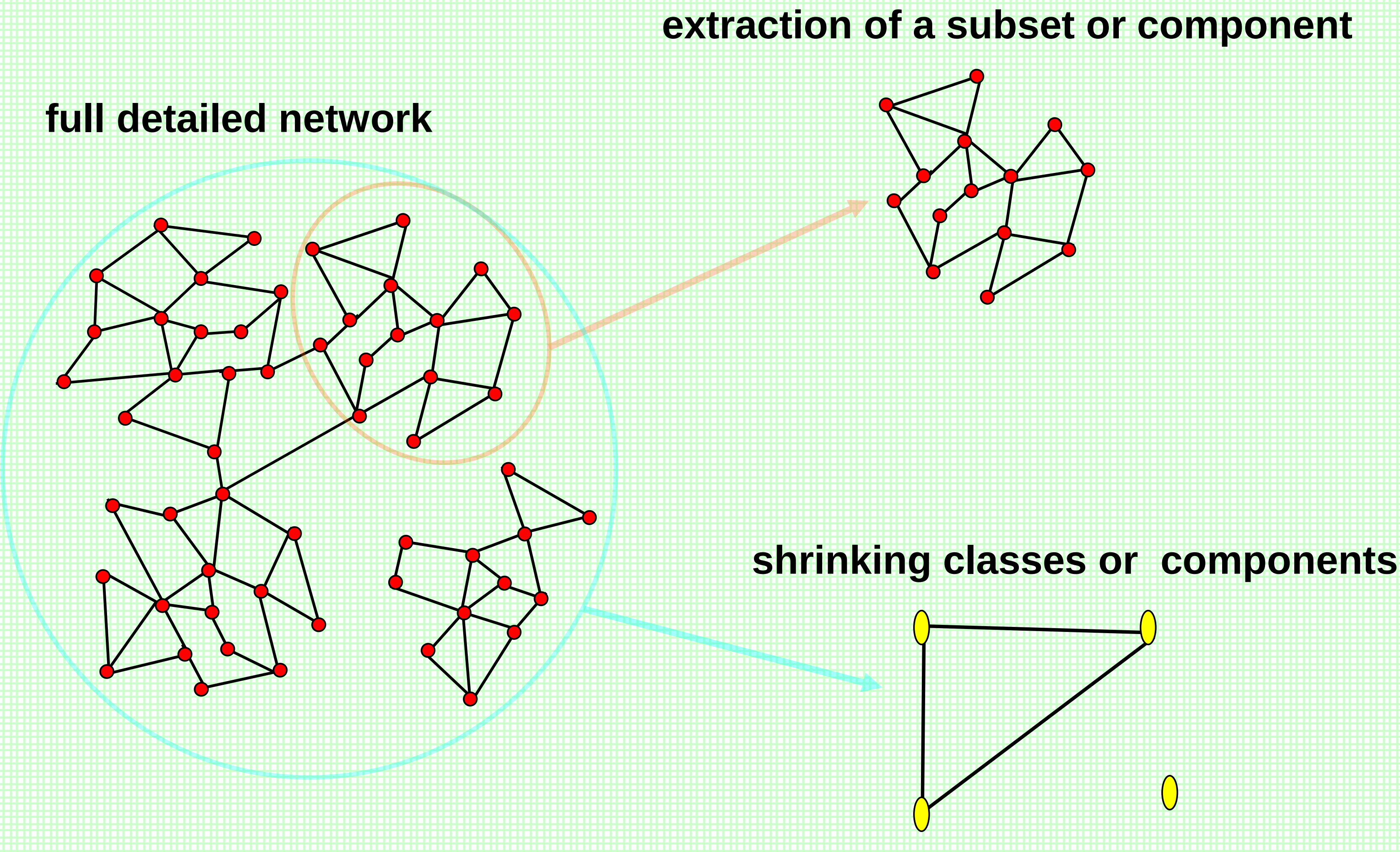
• two sets of vertices
• vertices can only relate to member of other set

movies, feed suppliers

actors, farms



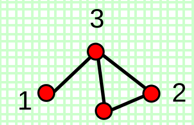
Definitions, Concepts & Tools cont.

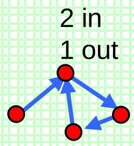


Definitions, Concepts & Tools cont.

vertex degree

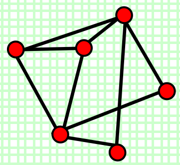
- **# of incident lines (network avg.)**
- **in-degree, out-degree**
- **adjacent or neighbouring nodes (by connection not distance)**

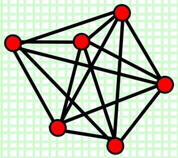




density or clustering coefficient

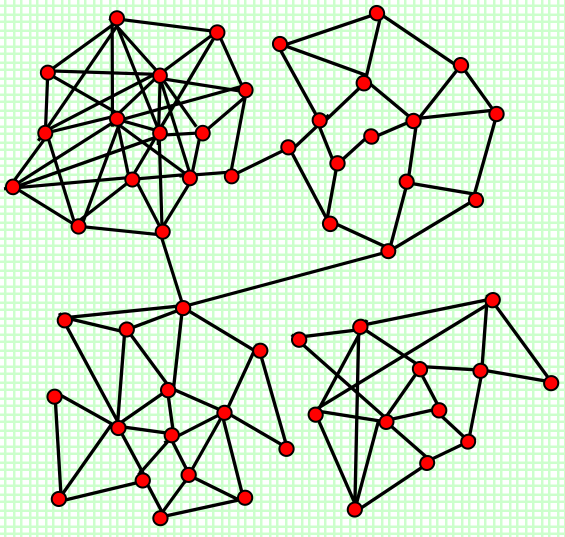
- **# of lines as a proportion of max. (9/15)= 0.6**
- **complete network with maximum density k(k-1)/2 = 15**





network components **(sub-networks)**

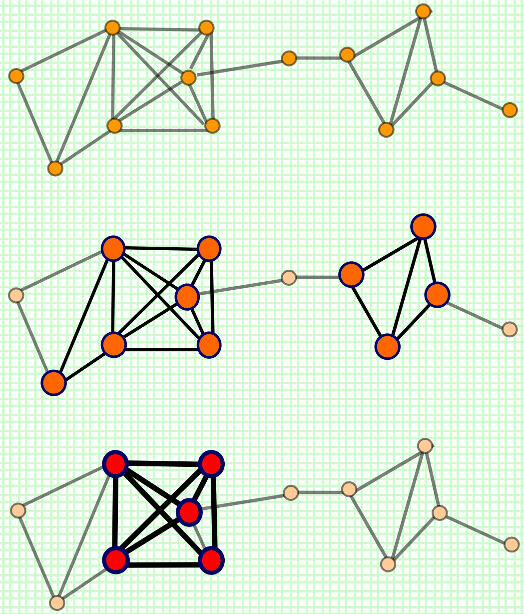
- **strongly and weakly connected components**
- **clusters of connectivity (not spatial)**



Definitions, Concepts & Tools cont.

k-core

maximal sub-network in which each vertex has at least degree k within the sub-network

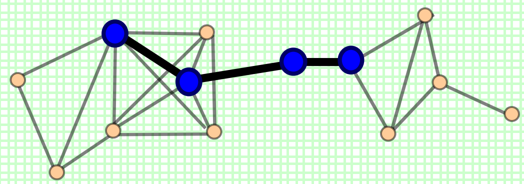


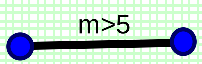
3-core

4-core

m-slice

maximal sub-network containing nodes with lines with multiplicity of at least m (e.g. 5)





Definitions, Concepts & Tools cont.

semi-walk

sequence of lines between two nodes that do not have to follow direction & can repeat nodes

walk

sequence of arcs between two nodes that follow arc directions, but can repeat nodes

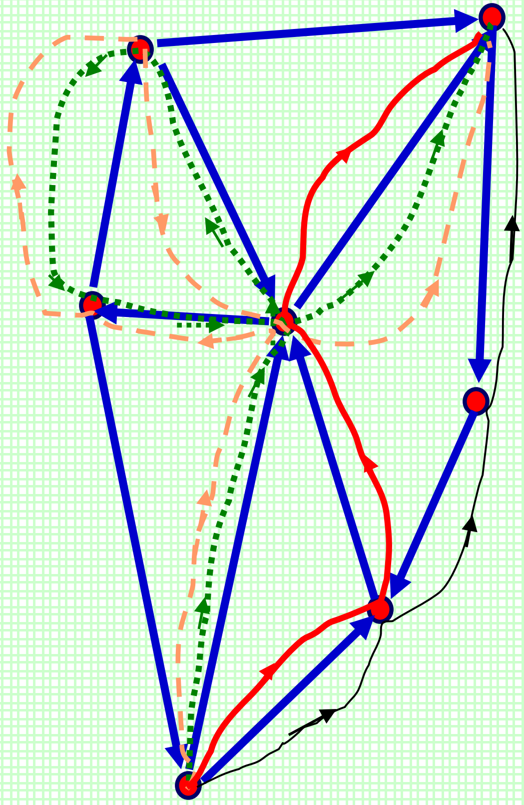
semi-path

semi-walk (not constrained by direction) but cannot repeat nodes

path

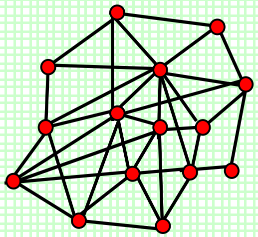
directed walk without repeating nodes (must follow arc directions & not repeat nodes)

Consider number paths and number of steps in path lengths, (min, mode, max)



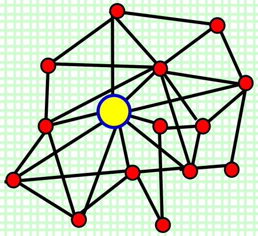
Definitions, concepts & Tools cont.

distance
number of line steps between two nodes
(e.g. avg. dist. vs. total diameter)

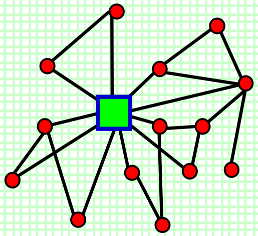


"small world"
Even in large networks the distance (number of steps) between any two nodes is usually relatively small
(approx log of N in random networks)

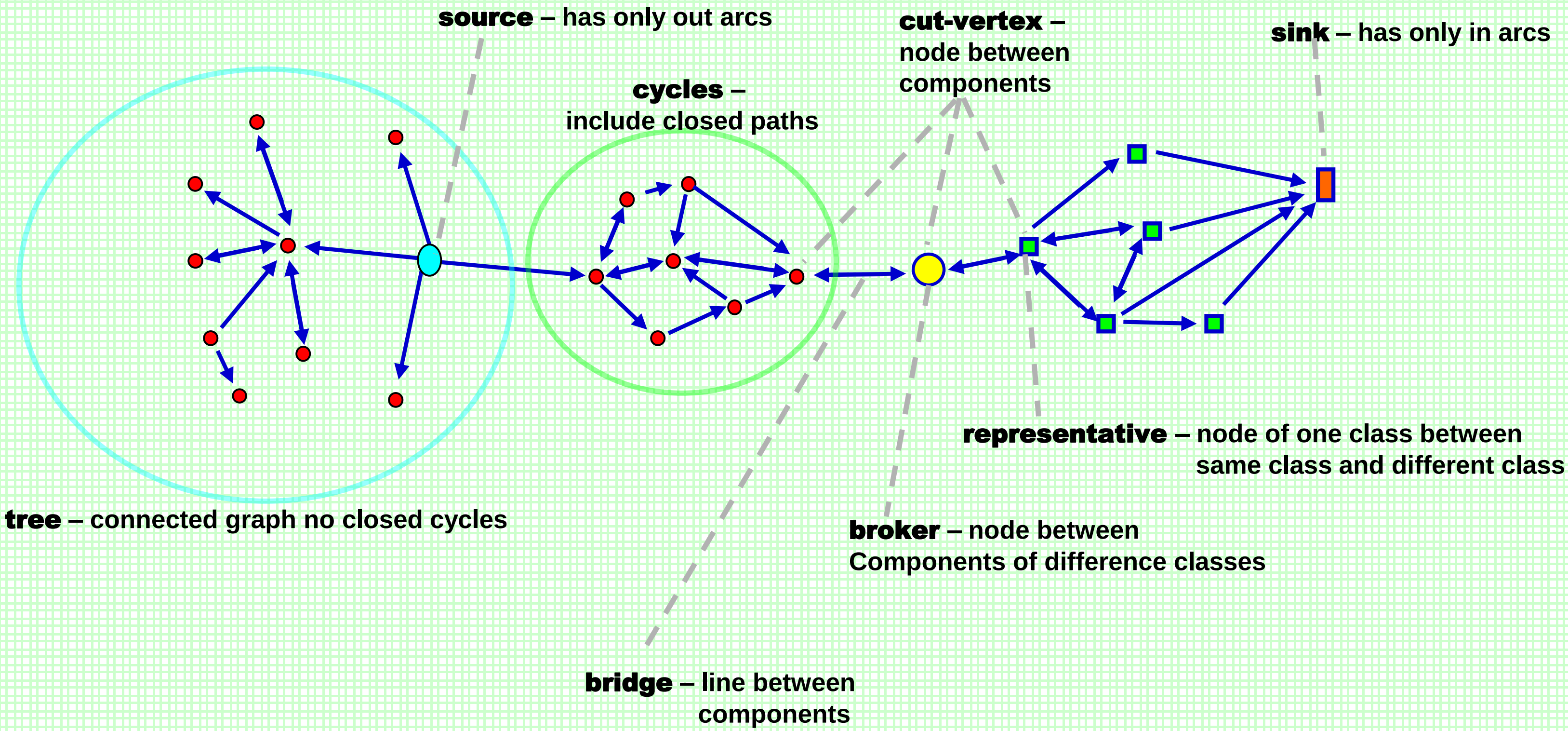
closeness centrality
of a vertex - refers to distance to other nodes as # of other nodes / sum of all distances form it to all others
i.e. shorter distance from this node to most other nodes than between most other pairs (e.g. 2 steps)



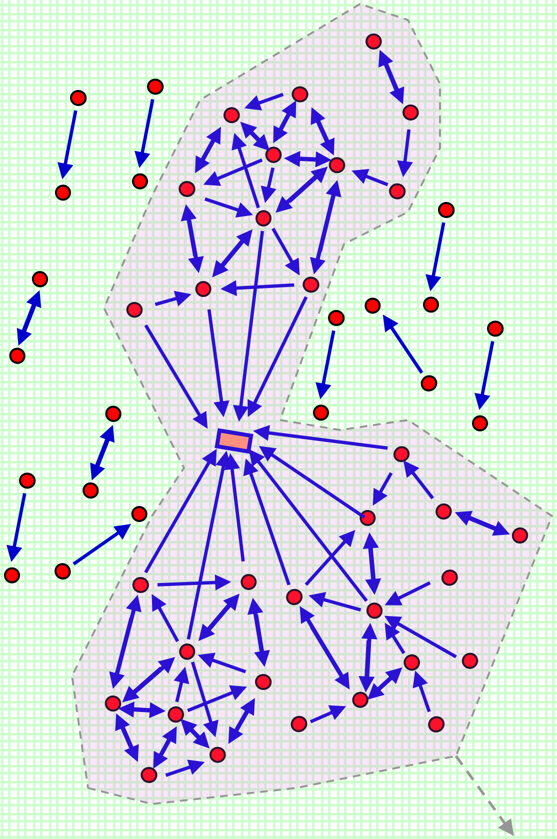
betweeness centrality
of a vertex – refers to proportion of shortest paths (geodesics) that include that node i.e. many paths between most pairs must go through this node

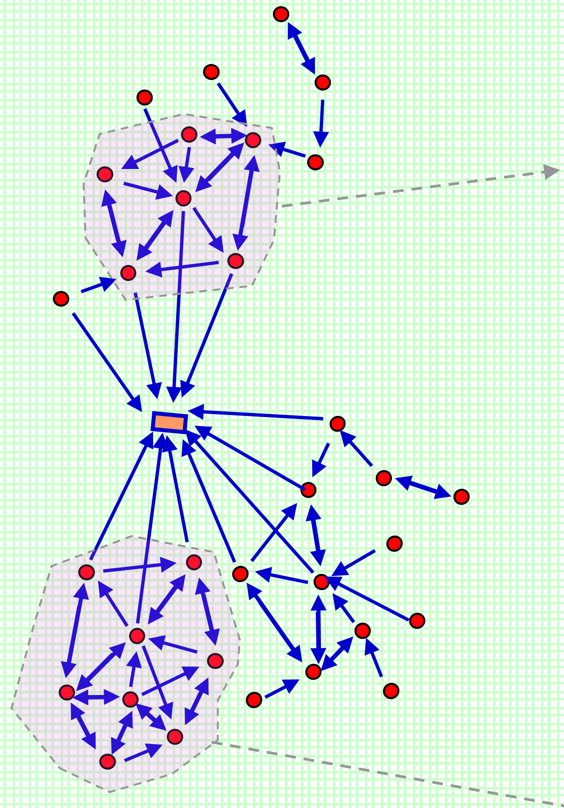


Definitions, Concepts & Tools cont.

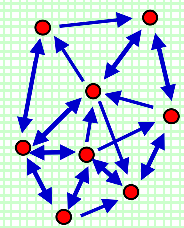


Definitions, concepts & Tools cont.





Strong component (not largest)
- connected, direction matters
- eg. direct contact



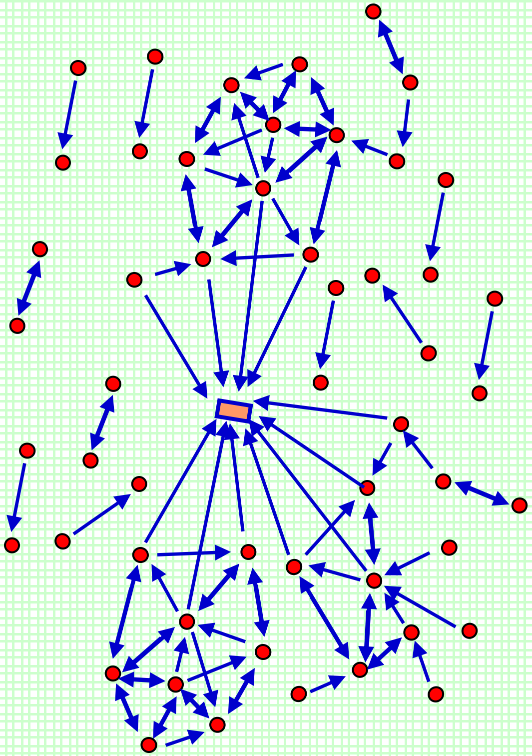
Full network

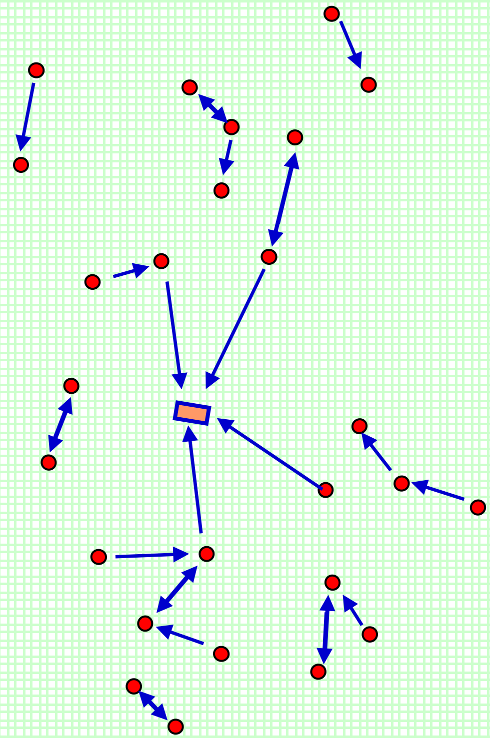
Largest weak component
- connected, direction not matter
- eg. indirect contact

Largest strong component
- connected, direction matters
- eg. direct contact

Definitions, concepts & Tools cont.

Network characteristics may be vastly different depending on duration of time observed
- What duration is relevant to the concern of interest ?





e.g. 6 mo network of contacts vs. 1 month network

Random vs. Scale Free Networks

Degree distribution
- distribution of the number of degrees (linkages) per node

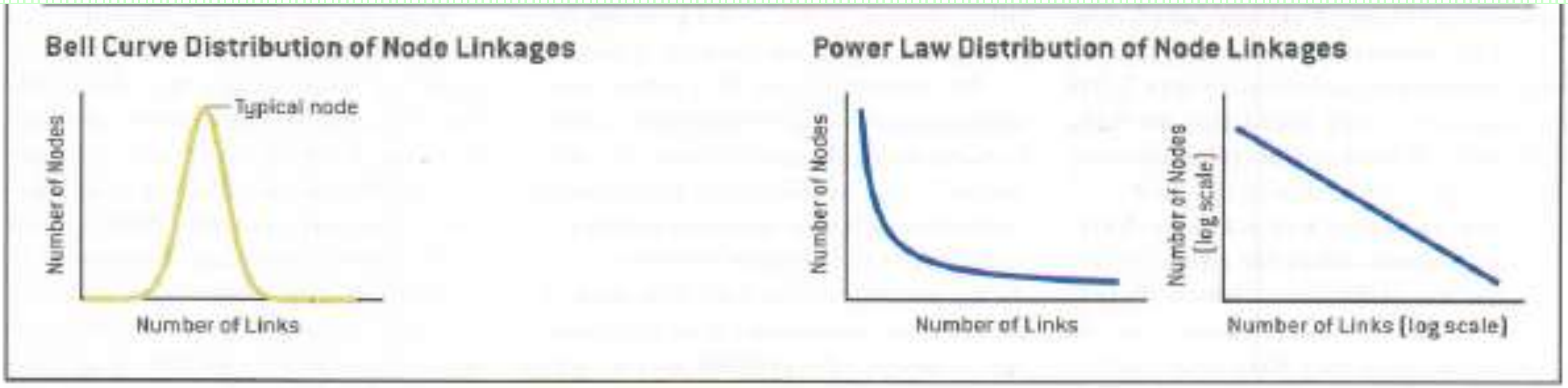
Random Networks (Erdös & Rényi 1959)
- nodes have roughly the same degrees… close to the average number of linkages
- random distribution
- probability of a node having exactly k degrees follows a Poisson with a good Binomial approximation

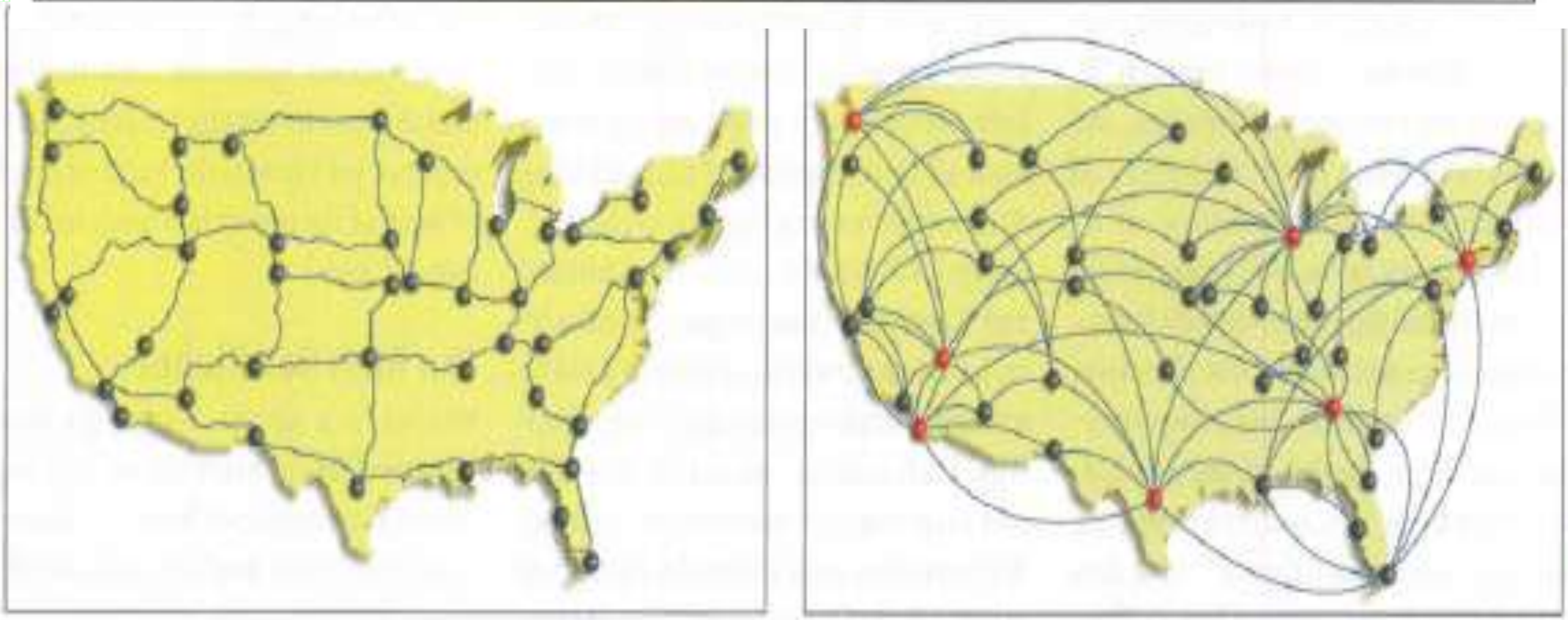
Scale-Free Networks (Barabási & Albert 1999)
- wide variability in number of degrees, some very high "off-the-scale"
- fallows power law distribution which is free of scale $p(k) \sim k^{-g}$ (often $2<g<3$)
- e.g.
  most nodes have 3 to 8 links, but a few hubs with thousands of links
  most people 5.25 to 6.25 feet tall but a few over 500 feet tall !!

Random vs. Scale Free Networks cont.

Random e.g. road network

Scale Free e.g. air network
(e.g. hubs Denver, Atlanta, Chicago)





Barabasi & Bonabeau Sci Am 2003

Random vs. Scale Free Networks cont.

Examples of Scale-Free Networks

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

Barabasi & Bonabeau Sci Am 2003

Consider nature's chaos, fractals, complexity, apparent ordered randomness…
scale-free networks that grow out of preferential attachments describe a lot of stuff !!!

Implications for Spread & Control

Network characteristics
e.g. connectedness, clustering, degree distribution etc.

Diffusion
Network characteristics greatly influences the rate and extent of diffusion (spread) through a network

Communications and marketing networks
we want extensive, rapid diffusion

vs.

Disease spread networks
we want no, poor, or slow diffusion (spread)
e.g. NAADSM contact frequencies & movement controls
or peace-time re-design of livestock networks

Implications for Spread & Control

In Social Network Parlance

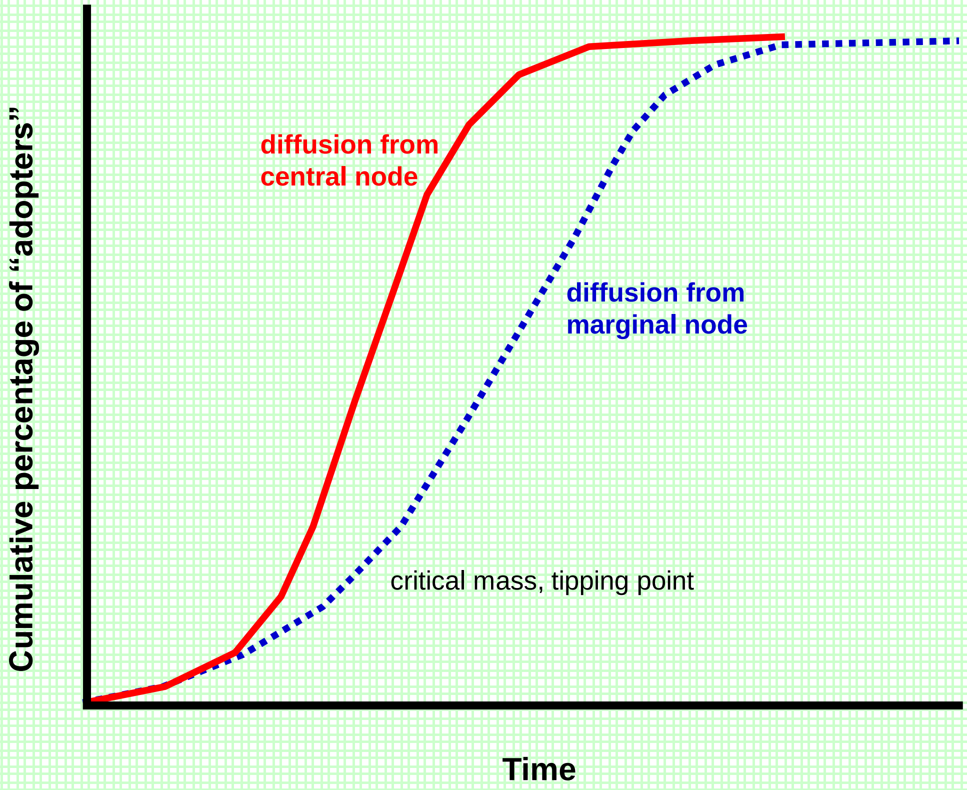
adoption rate (incidence)

exposure (NAADSM contct. feq)

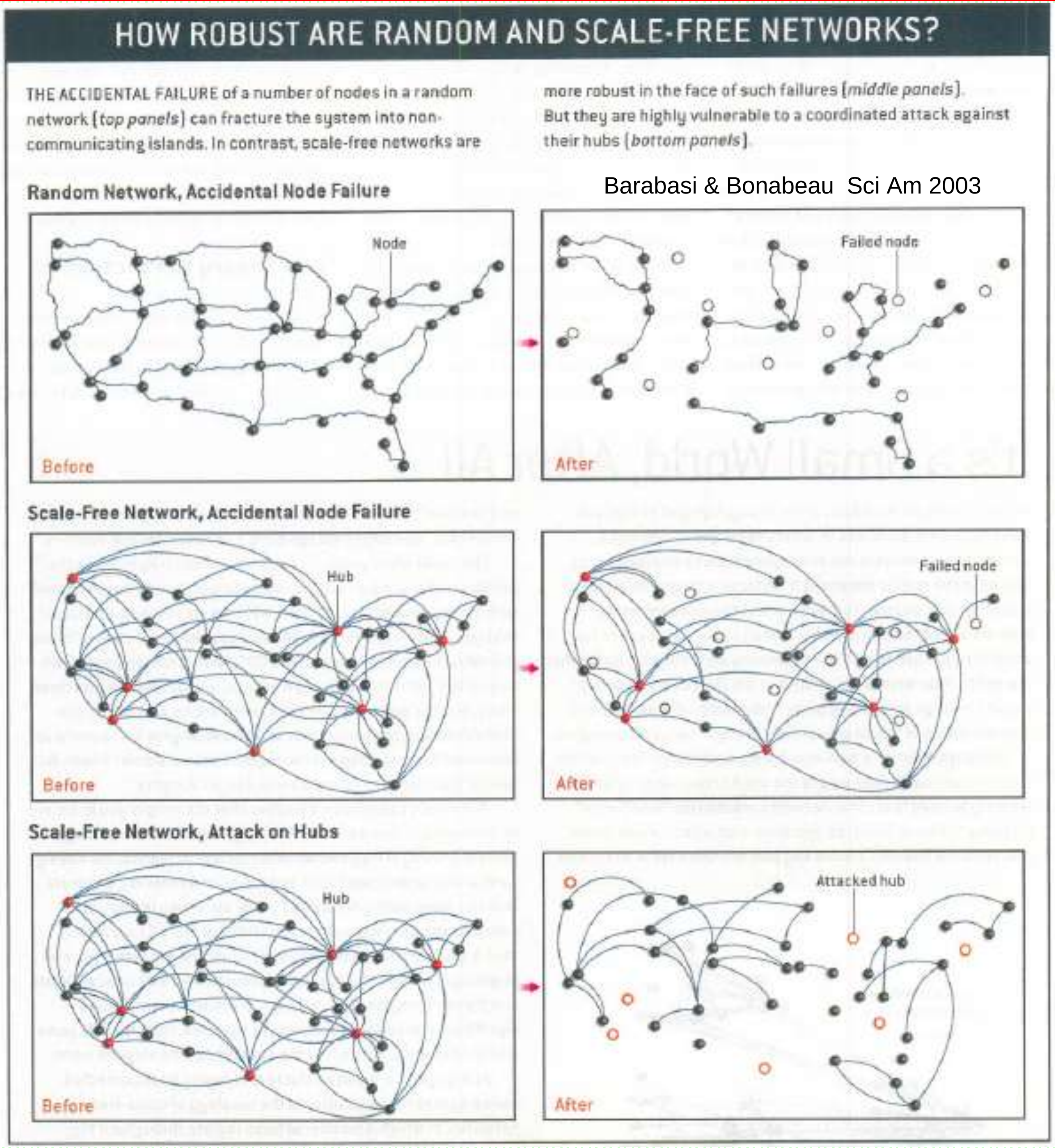
threshold (NAADSM p infect.)

threshold lag (latent)

critical mass (R >1, Sum p > 1)

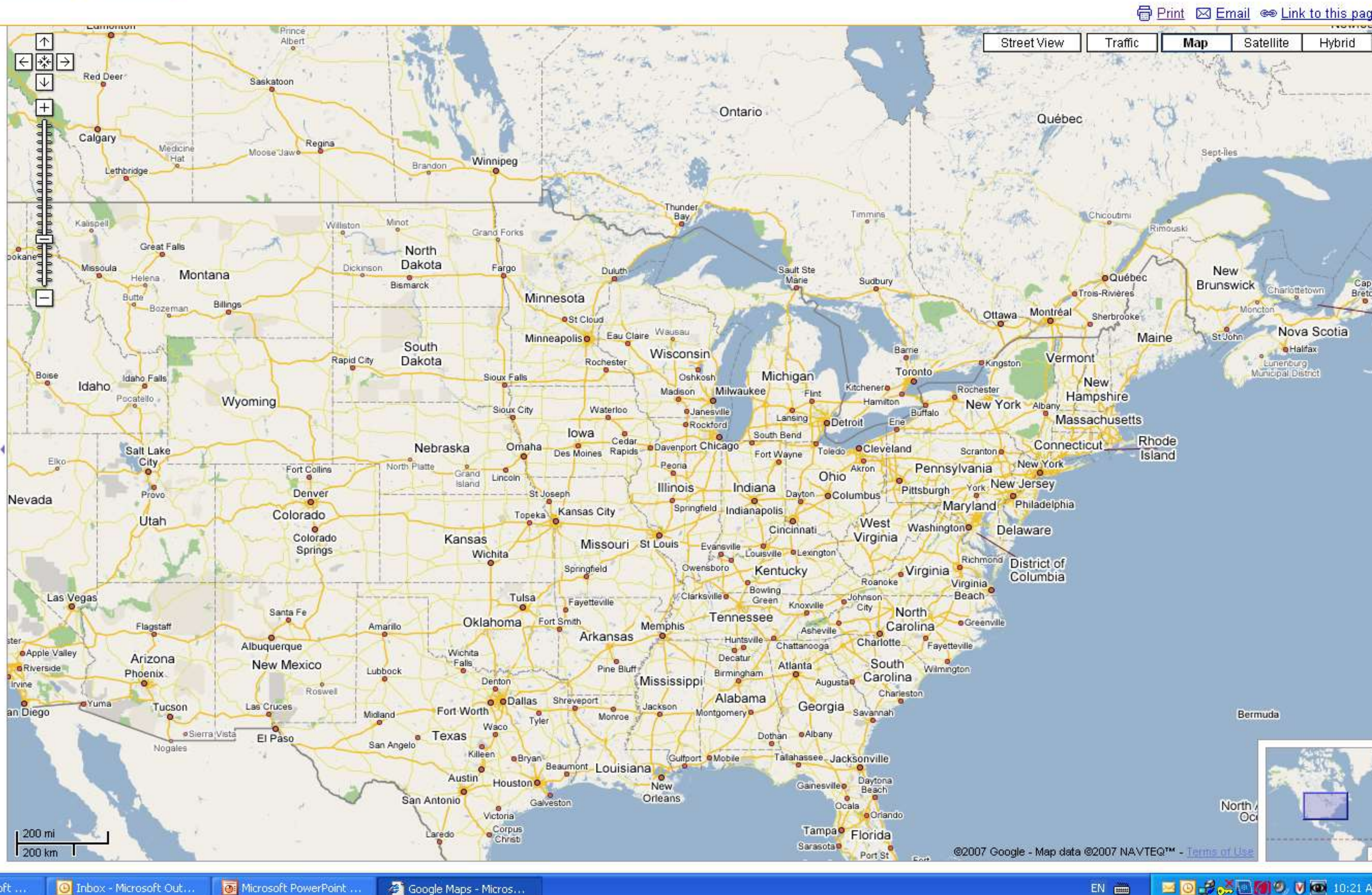


Implications for Spread & Control cont.



This SFN characteristic is both useful and scary !!!

Implications for Spread & Control cont.



Livestock Examples of Scale-Free Networks

Network analyses of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread.
Bigras-Poulin et al, Pev Vet Med, 76:11-39, 2006

- clustering coefficient of 0.52 inward and 0.02 outward

- 130,265 arcs (at least one cattle movement) among 19,805 connected nodes (premises) for an average degree connectance of 6.58 arcs per node (6 mo)

- distribution of degrees followed power-law with in-degree power of -2 and out-degree power of -1.46 (i.e. scale-free)

- most non-abattoir movements were between two farms, involved different animals, only a few animals per movement, and occurred locally geographically.

- dealers or live markets involved in 3% of movements, but those premises had higher degrees, larger premises specific sub-networks, more steps (network "distances") and greater geographic distances in paths to and from them.

- map of nodes with color coding of in and out degree could provide a basic risk map

- understanding livestock network and flow is important to understanding disease spread and control

Livestock Examples of Scale-Free Networks

Network analyses of Danish cattle cont.

Bigras-Poulin et al, Pev Vet Med, 76:11-39, 2006

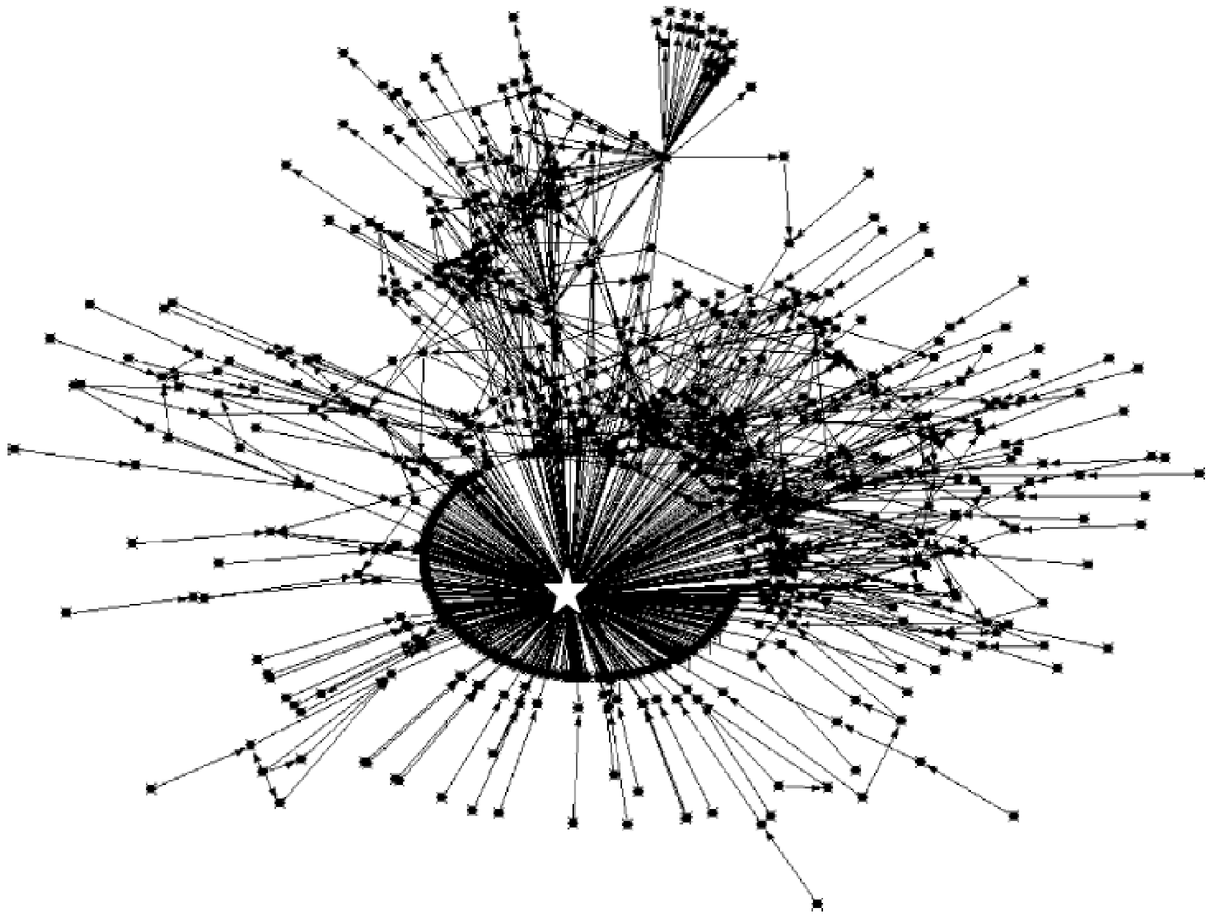


Fig. 7. A market (shown by the star) with movements of animals that reached other premises in up to three links inward and outward of the specific market, during the period extending from November 1, 2002 to April 30, 2003.

Livestock Examples of Scale-Free Networks

Relationship of trade patterns of the Danish swine industry animal movements network to potential disease spread.
Bigras-Poulin et al, Pev Vet Med, 80:143-165, 2007

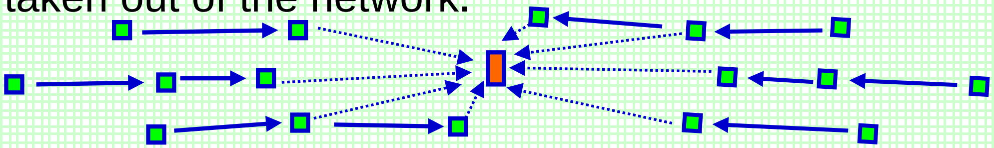
- assessed cluster coefficients, degree distribution, connectance, adjacentcy, reachability, path-lengths, cycles and many other characteristics.

- overall clustering coefficient of 0.064 inward and 0.005 outward

- overall 43,940 arcs (at least one swine shipment) among 14,548 connected nodes (premises) for an average degree connectance of 3.02 arcs per node (6 mo)

- distribution of degrees followed power-law with in-degree power of -0.57 and out-degree power of -2.30 (scale-free)

- But, if exclude abattoir nodes, there were 11,217 arcs among 6666 connected nodes (farms and markets) in the sub-network for an average degree of 1.68 links per node; Note that a large number of nodes (7882) became disconnected when slaughter related movements were taken out of the network.



- Overall, avg. path length 3.19 steps, max. path length 8 steps (i.e. Small World)

Livestock Examples of Scale-Free Networks

Network analysis of Danish swine cont.

Bigras-Poulin et al, Pev Vet Med, 80:143-165, 2007

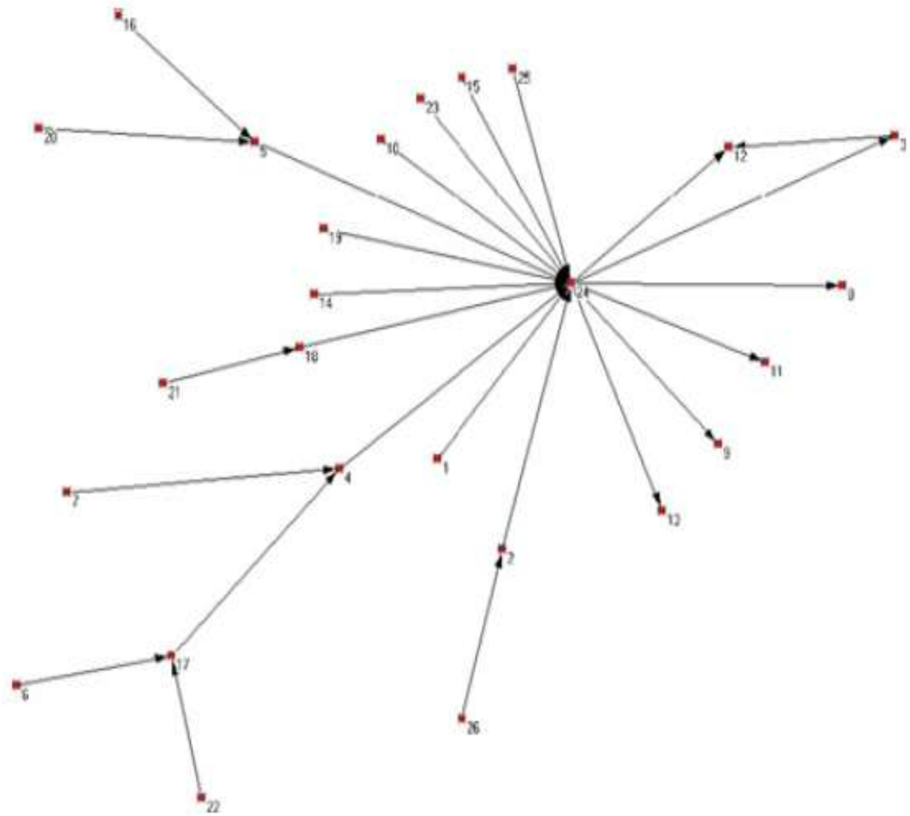


Fig. 4. Premise specific network of a Danish slaughter pig farm (indicated by 24), with movements of animals that reached 19 premises (to) and 6 premises reached by the farm (from), in up to 3 steps, during the period extending from November 1, 2002 to April 30, 2003.

Livestock Examples of Scale-Free Networks

Network analysis of Danish swine cont.

Bigras-Poulin et al, Pev Vet Med, 80:143-165, 2007

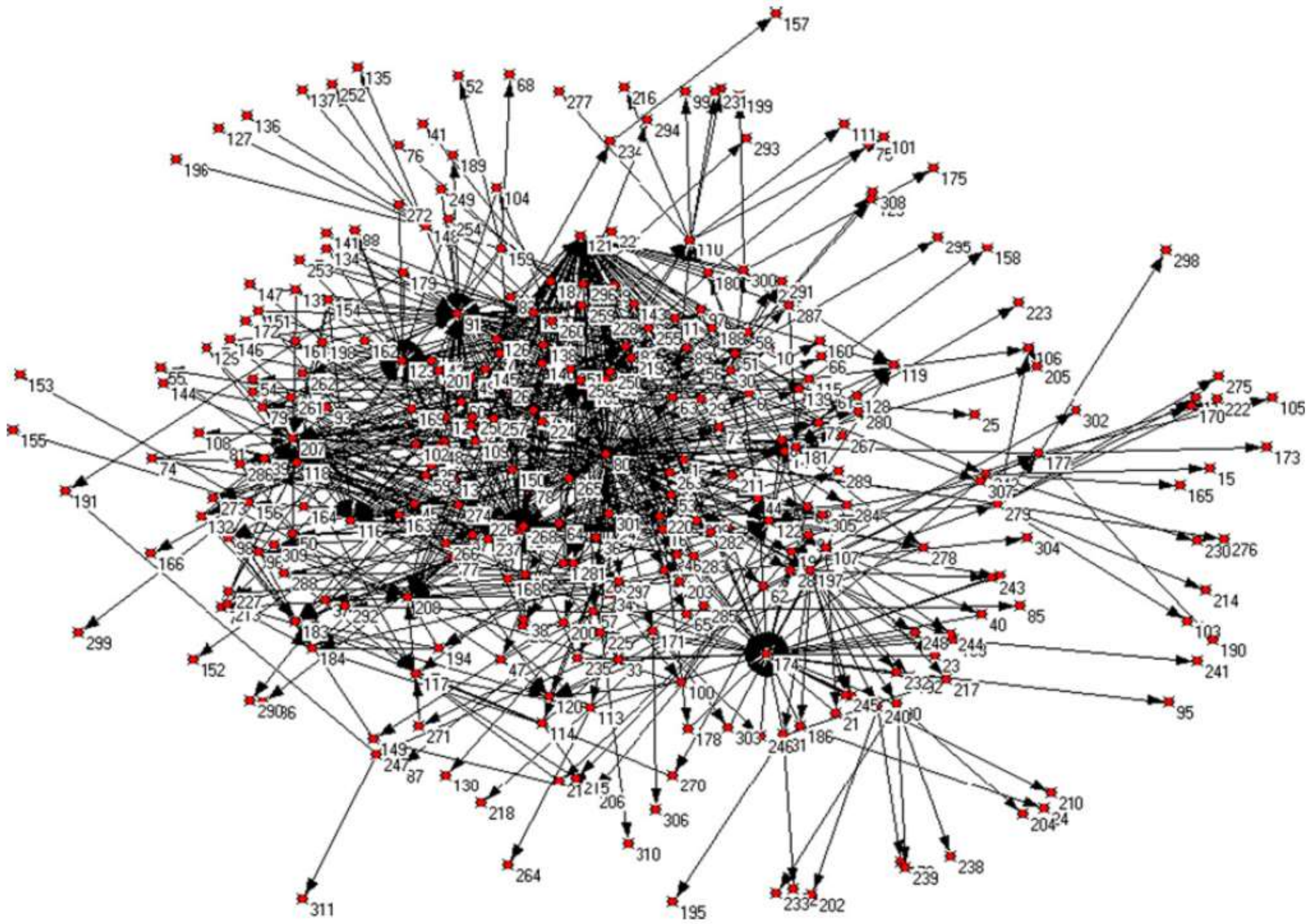


Fig. 5. Premise specific network of a Danish breeder farm (indicated by 80, center) with movements of animals that reached 0 premise (to) and 269 premises reached by the farm (from), in up to 3 steps, during the period extending from November 1, 2002 to April 30, 2003.

Livestock Examples of Scale-Free Networks

Network analyses of Canadian dairy cattle movements, in Ontario, 2004-2006 Dubé et al, (work in progress)

- 9890 arcs (cattle movements) among 3212 connected nodes (premises) for an average of 3.1 arcs per node (DHI data over three years)

- Overall clustering coefficient of 0.075 and fragmentation of 0.717 (meaning 71.7% of pairs of nodes were unreachable)

- distribution of degrees followed power-law with
    - in-degree power of 1.9 and out-degree power of 1.6

- path lengths 7.07

- Compared to a random graph with the same number of nodes and density, the Ontario dairy cattle movement network displays small-world properties (clustering)

- characteristics of 3-yr-network very different from those of 1-mo-network

- classical network analyses of strong and weak components may be less useful

- **understanding livestock network and flow pertinent to the biology of the disease of concern is important to understanding disease spread and control**

Livestock Examples of Scale-Free Networks

Network analyses of Canadian dairy cattle cont.

Dubé et al, (work in progress)

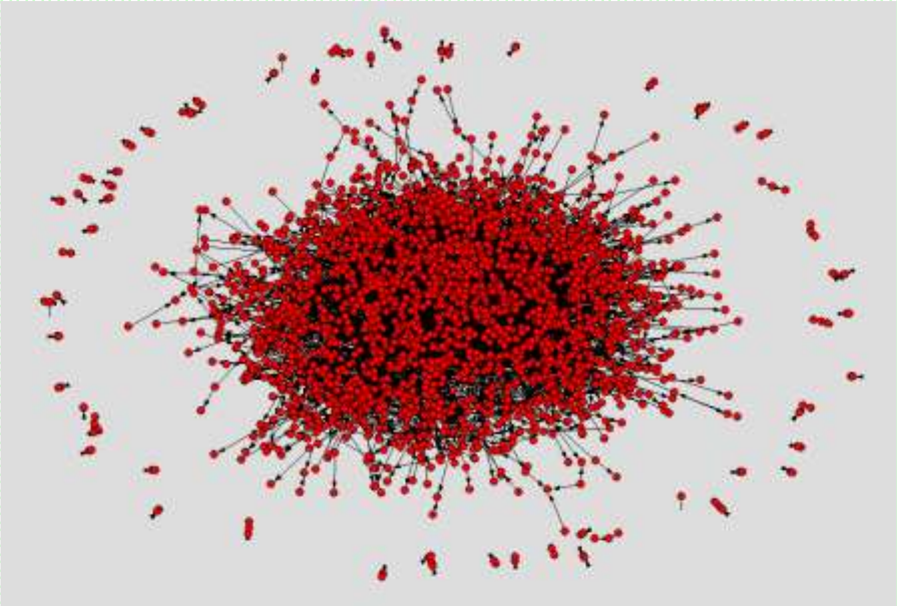


Figure 1. Overall network of adult milking cow movements among Ontario dairy farms enrolled on the Dairy Herd Improvement Plan (DHI) 2004-2006.

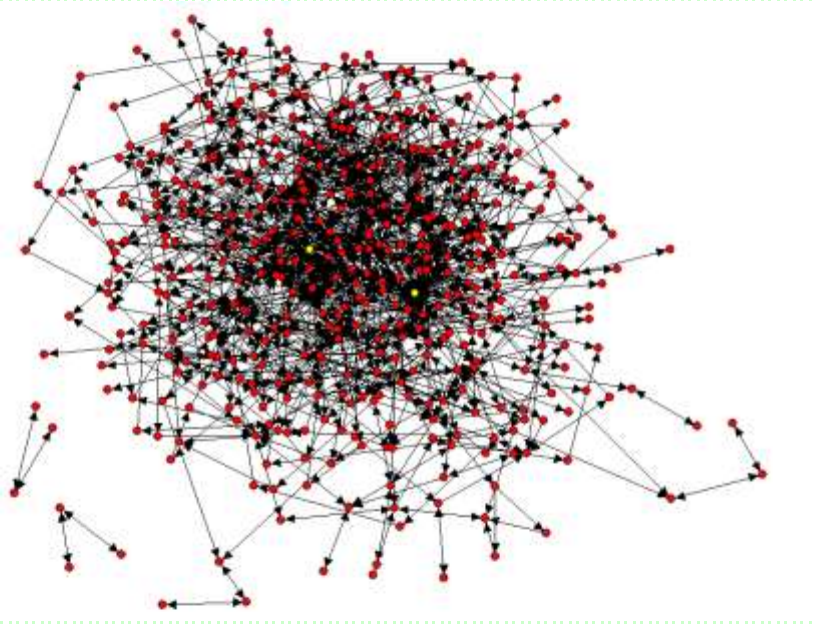


Figure 2. Giant strong component formed by the movements of adult milking cattle among DHI herds in years 2004-2006. The two nodes in yellow are highly connected nodes acting as hubs in the network of farm to farm movements.

NAADSM and Networks

- **NADDSM is NOT designed to analyze networks (i.e. not like Pajek or UCInet)**
- **Bigras-Poulin (PVM 2007) specifically referenced earlier version of NAADSM Schoenbaum & Disney (PVM 2003) (and mathematical models), as modeling disease spread as if the network of contacts was random; suggested this is reasonable when used to evaluate different disease control strategies in general as a preparedness tool, but NOT a good predictive tool**
- **NAADSM v3 does NOT facilitate modeling of explicit networks of specific links (edges or arcs) between specific premises (nodes) (we don't have the data anyway !!)**
- **BUT, through extensive use of NAADSM unit-type and specific type-to-type contact parameters, one can partially simulate (not analyse) non-random networks.**
- **Setting 0 contacts between two different production types has the same effect as removing links between those types of nodes in the network,**
- **Setting increased frequency of direct and indirect contacts from one type to another is analogous to increasing the out-arcs from those source unit types**
- **If have relatively few units of one type (eg sales yard) and have many units of many other types capable of being sources for direct or indirect contact with sales-yards, then even at low frequency per source units, it still has the effect of generating many in-arcs to those few sales-yards (until they are detected as positive or trace)**

The Future of Network Analyses & NADDSM

- In general, network analyses is now and will become increasingly important for all sorts of value-added-business-reasons, far beyond disease control

- NAADSM team wants to somehow (eventually) incorporate more accurate networks into NAADSM (v-4 v-5 ?)

- Not yet clear exactly how we are going to do that, but it is important to try

- Want to learn more about networks and work with current NAADSM to simulate non-random networks through extensive use of unit-types and type-to-type specific parameters (tedious)

- Want to do the above and then analyze NAADSM outputs in "network-analyses-software"

- e.g. load NAADSM contact tables into Pajek for analyses of the networks that have been created by NAADSM…see how close their network-descriptive-parameters are to known networks (e.g. Ontario cattle movements)

- added value to premises ID and movement recording for tracing (we will have to do it eventually, why not get on with it now ?)

Back to Session Outline

- ✓ Some examples of networks
- ✓ Definitions & tools to depict, describe & analyze networks
- ✓ Random vs. scale-free networks
- ✓ Implications for spread & control
- ✓ Livestock examples of scale-free networks
- ✓ NADDSM modeling & networks (now and future)

QUESTIONS ?