Class 15: Search and Navigation on Social and other Networks: Affiliation Hierarchies

Christopher L. Magee

ESD 342

April 1, 2010

Massachusetts Institute of Technology
Engineering Systems Division

Lecture 15 Outline

- Introductory Remarks
- Search and Navigation, search (briefly)
- Navigation
  - Milgram's experiment and critiques
  - Small World and predecessor models
  - Kleinberg's first model
  - The influence of structure and Kleinberg's second model (and the Watts, Dodds and Newman model)
- Modeling Overview
  - Materials metaphor
  - Search/navigation as case study of model evolution
  - Modeling limitations and benefits

Search and Navigation

- Search
  - "To look over carefully in order to find something, to explore", "to make an effort to find something" seek, hunt, quest.

- Navigate
  - "To plan, record and control the position of.." "to follow a planned Course" or "*to make one's way*"

Search and Navigation

- Search
  - "To look over carefully in order to find something, to explore", "to make an effort to find something" seek, hunt, quest.
  - Network literature: **"to find the node containing information that is desired"**
- Navigate
  - "To plan, record and control the position of.." "to follow a planned Course" or "*to make one's way*"
  - Network literature: "**to get from one to another specific node by a(the) short(est) path using *only local information***"

Network Search I

- Exhaustive WWW Search
  - Catalog (while "crawling") the network and create a map (local index) of the ***entire network***
  - Use information in nodes to select relevant web pages
  - Rank nodes for significance using the ***link*** information
    - Eigenvector Centrality (Brin and Page)
    - Each node has a weight x_i that is defined to be proportional to the weights of all nodes that point to i
    - And $x_i = \lambda^{-1} \sum_j A_{ij} x_j$
    - And then Ax = λx
    - Thus the weights are an eigenvector of the adjacency matrix (A) with eigenvalue λ

Kleinberg has considered a more sophisticated version with weights for *pointing* hubs and *receiving* hubs

Network Search II

- Guided Search- databases or the web for unmapped elements
  - Web “spiders”
  - Queries passed (if not answered) to highest k node of neighbors. If some nearest neighbor information is also stored this is a valuable approach for peer-to-peer systems

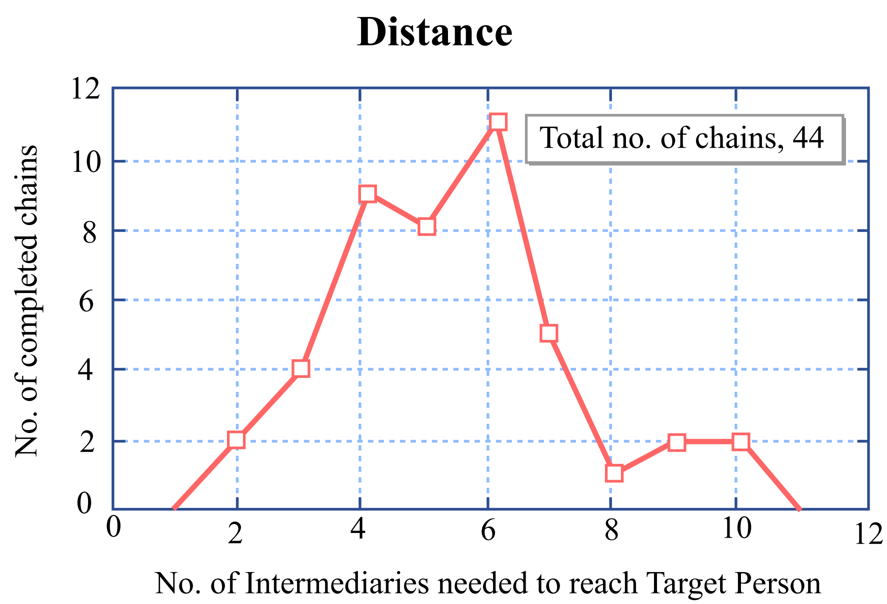
Network Navigation: Milgram's experiment

- The unpublished (but widely circulated) paper of Kochen and Pool using simple random graph models indicated the possibility of short paths through social networks. This instigated the famous social scientist Stanley Milgram to try an experiment
- "route" letter to person XXX who is a stockbroker living in Sharon, MA who works in Boston.
- The letters can ***only*** be sent to someone who the recipient knows on a ***first name*** basis but in a way to get "closer" to person XXX. Participants also were asked to record and send along the routing information

Milgram’s experiment

- “route” letter to person XXX who is a stockbroker living in Sharon, MA who works in Boston.
- The letters can ***only*** be sent to someone who the recipient knows on a ***first name*** basis but in a way to get closer” to person XXX.
- Some guesses that it would take hundreds of steps were refuted by results that “showed” it took (***actually can take***) much less

Six-degrees of separation



In the Nebraska Study the chains varied from two to 10 intermediate acquaintances with the median at five.

Image by MIT OpenCourseWare.

Milgram, *Psych Today* **2**, 60 (1967)

Milgram’s experiment II

- “route” letter to person XXX who is a stockbroker living in Sharon, MA who works in Boston.
- The letters can ***only*** be sent to someone who the recipient knows on a ***first name*** basis but in a way to get closer” to person XXX.
- Some guesses that it would take hundreds of steps were refuted by results that “showed” it took (actually can take) much less
- A play was written and coined the phrase “*six degrees of separation*” as its Title and Milgram’s result became something “everyone knows”.
  - ***Everyone*** is separated by only six removes from ***everyone*** else on the planet!
- But What did Milgram really show?

What did Milgram really show?

- Of 300 letters in original experiment, only 96 (random Nebraska) sampled tested the "***everyone***" part of what ***"everyone***" knows
- Only 18 of these were ***ever returned*** (the preceding graph contained very "non-random" Nebraska letters)
- Other trials that were random and tested the everyone basis had even ***smaller return rates*** than Milgram's initial experiment

- Issues
  - Is ***everyone*** really 6 or less steps from ***everyone***?

Lecture 15 Outline

- Introductory Remarks
- Search and Navigation, search (briefly)
- Navigation
  - Milgram's experiment and critiques
  - Small World and predecessor models
  - Kleinberg's first model
  - The influence of structure and Kleinberg's second model (and the Watts, Dodds and Newman model)
- Modeling Overview
  - Search/navigation as case study of model evolution
  - Modeling limitations and benefits

Poisson Random Graph

- Rapaport and later Erdos and Renyi and others such as Bollobas have studied a very simple model ***in some depth***. This is the one where each node in a network is connected with probability p to other nodes. Ensembles with variable numbers of links <k>

  are studied and the degree distribution is

$$p_k \cong \frac{<k>^k e^{-<k>}}{k!}$$

- The path length can be formally shown to be

$$l \cong \frac{\ln n}{\ln <k>}$$

  and is thus consistent with a "Small World"

- Clustering is simply equal to the random probability of a link between 2 nodes and is

$$C = <k>/n$$

Poisson Random Graph II

- It is generally stated that this model is nice for intuition but describes no real networks. It also provides a **benchmark**.
- If we look at a wide variety of "real world" graphs such as Table II from Newman

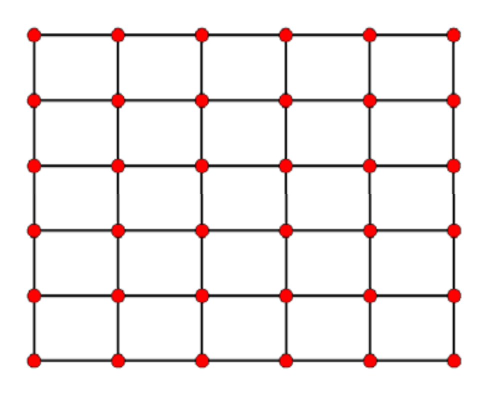
- What do we see?
- Path Length, *l*, is generally small (small worlds) and often approximately equal to path length for a Poisson random network
- Clustering is usually orders of magnitude higher than predicted by random networks for the large networks and is ~constant with n

Small World Problem as seen by Watts

Lattice

$$L(N) = N^{1/d}$$

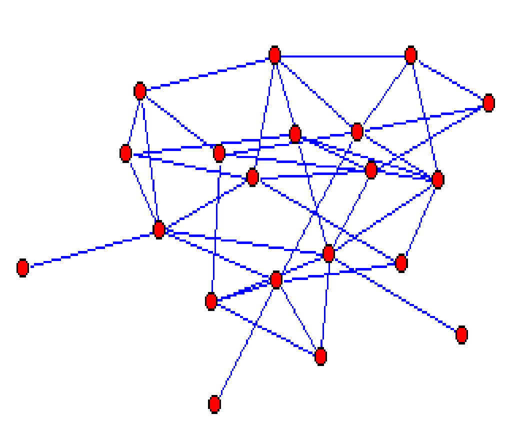
$$C(N) \approx const.$$



Random graph

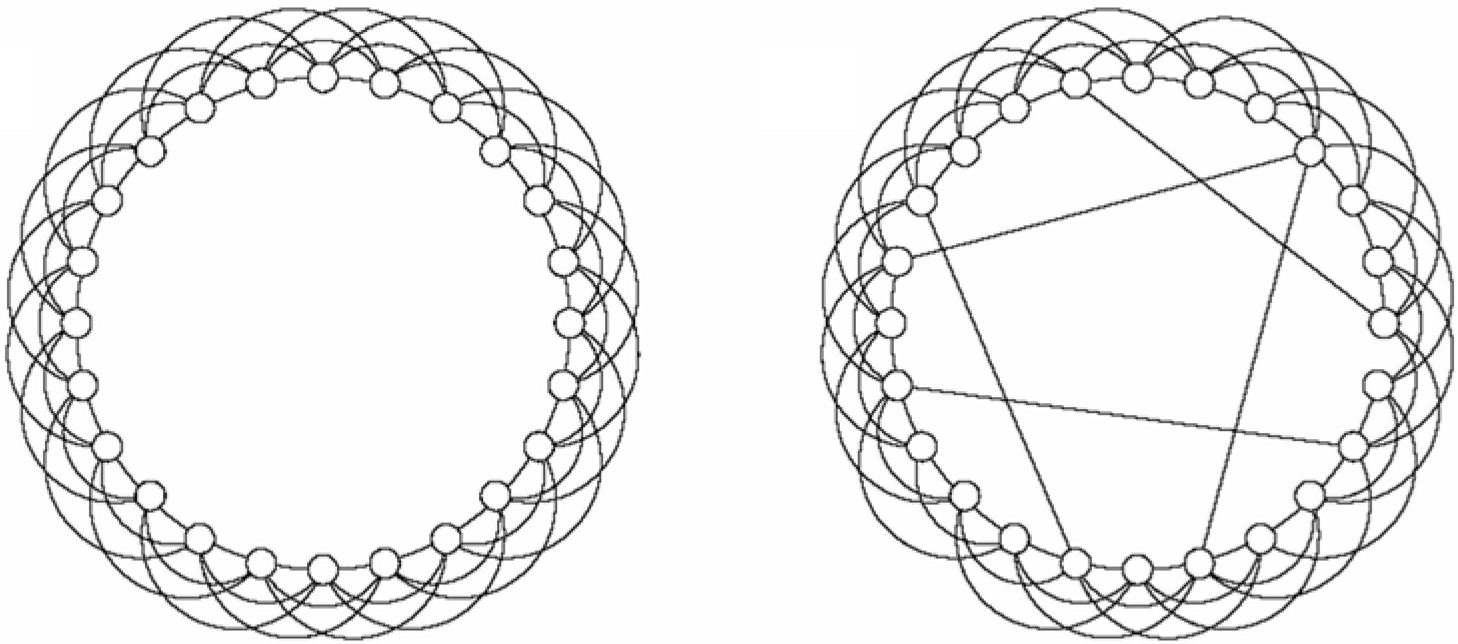
$$L(N) = \log N$$

$$C(N) \approx N^{-1}$$

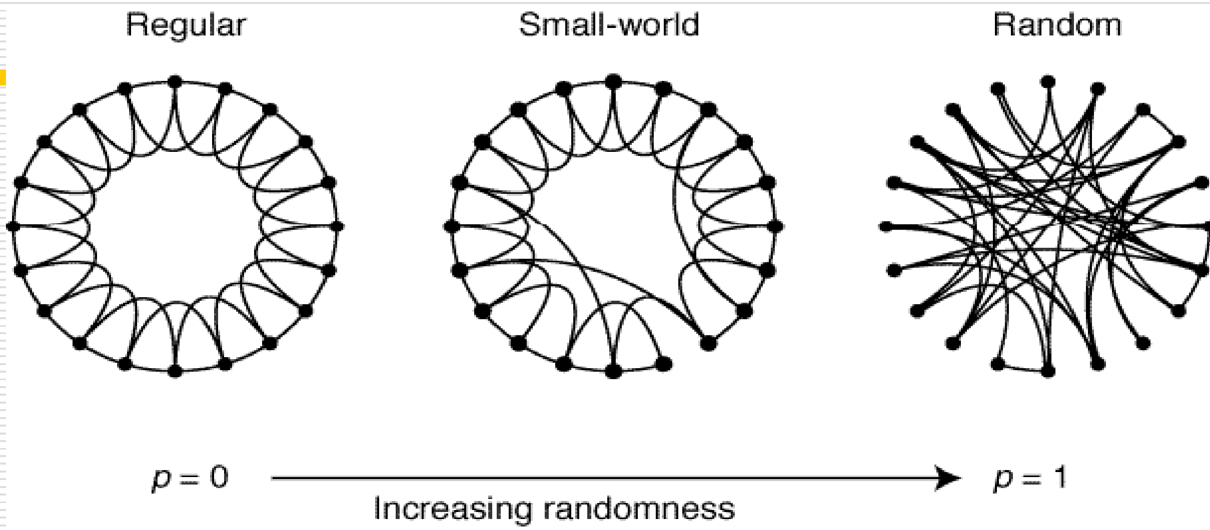


Small World Network Model (1D)

K is the number of nearest neighbors originally with links
(=3 below)



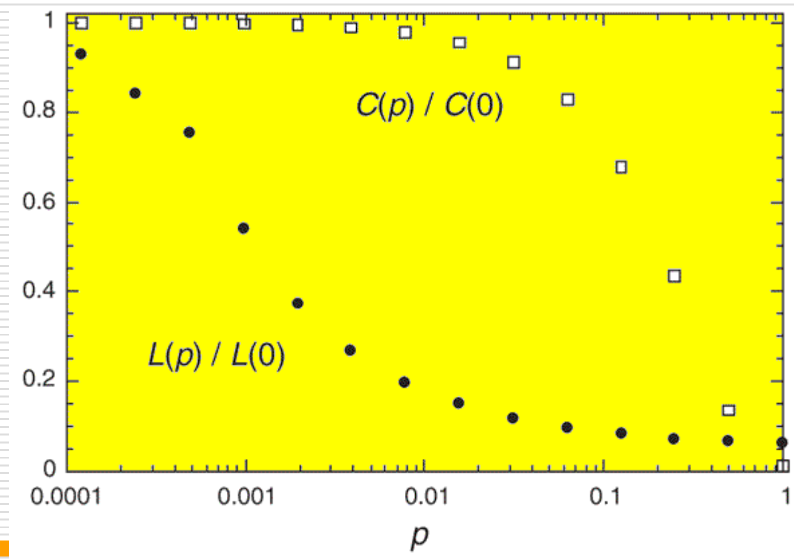
Small-world networks



- Large clustering coeff.
- Short typical path

Watts & Strogatz,

Nature **393**, 440 (1998)



N = 1000

Small World Clustering Estimation

- Watts and Strogatz got results from simulation
- Later Work by Barrett and Weigt on their model derived a clustering coefficient of

$$C = \frac{3(K-1)}{2(2K-1)}(1-p)^3$$

- An improved model by Newman and Watts and independently by Monasson gives for the clustering coefficient

$$C = \frac{3(K-1)}{2(2K-1)+4Kp(p+2)}$$

- These estimates are sufficiently high for real networks

Small World Model Path Lengths

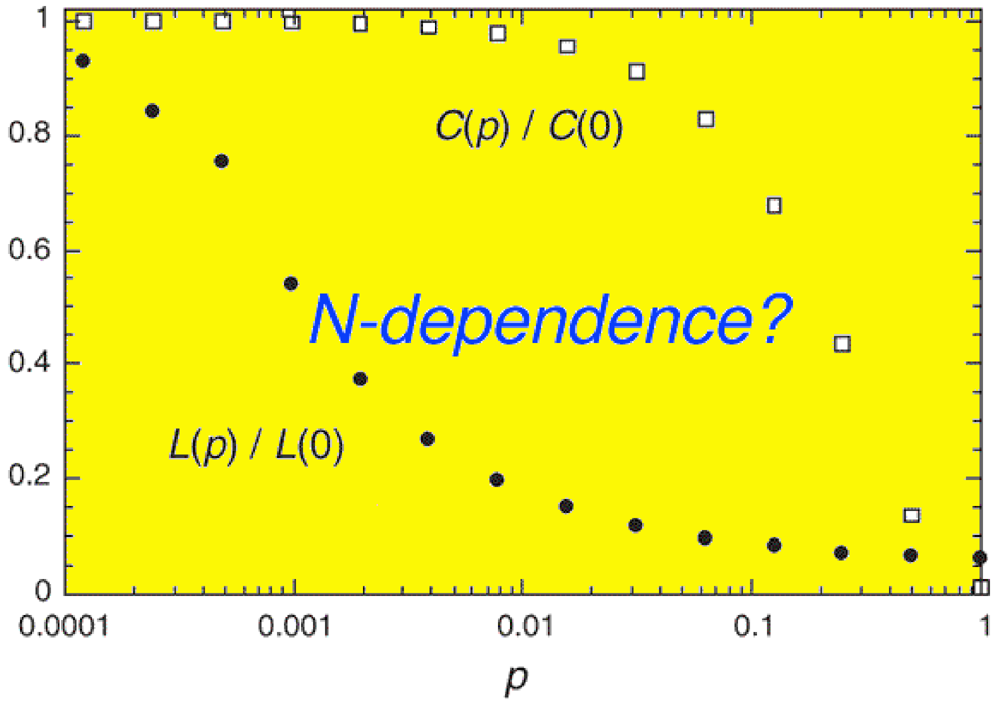
- Simulation based by Watts and Strogatz showed that path lengths were small and scaled with ln n

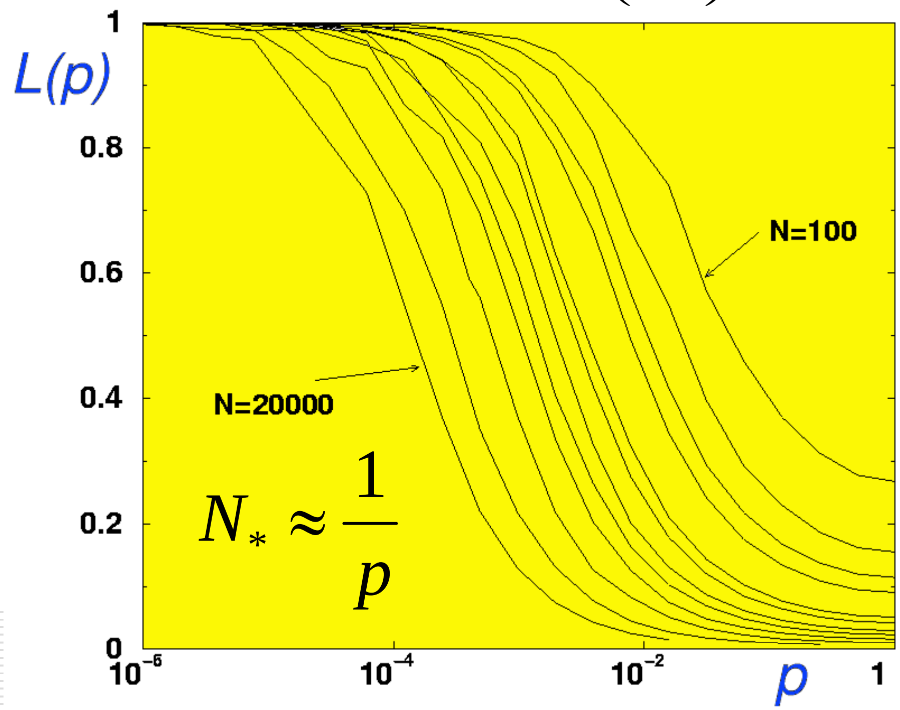
- No exact solution (yet) but Barthelemy and Amaral proposed a scaling relation that was later derived by Newman and Watts. It shows that the transition to “Small World Path Length Dependence” occurs at smaller p as n increases. Indeed, ***the number of shortcuts needed to give small world behavior is constant*** (for given K) ***as n increases***

$$l = \frac{n}{K} f(nKp)$$

Ubiquity of small-world networks

$$L(p,N) \sim N_* F\left(\frac{N}{N_*}\right)$$





Bertelemy and Amaral, Phys Rev Lett **83**, 3180 (1999)

Newman & Watts, *Phys Lett A* **263**, 341 (1999)

Barrat & Weigt, *Eur Phys J B* **13**, 547 (2000)

Small World Models

- Small world models thus
  - Show that it is relatively easy to have higher clustering and yet short paths. In large networks a few long paths is all that is needed- brain now understood this way as are some other large scale complex systems
- However, the specific models have only marginal connection to any real systems as they are stylistic and notional
- Small World Models have been relatively widely used as a "substrate" for studies of such as iterated games, epidemics. The rewiring approach has also proven useful even if the specific models are not real

Potential short paths

- There are almost surely *relatively* short paths between **any** two individuals

- The path length is *apparently* about that calculated for random networks:

$$l \cong \frac{\ln n}{\ln < k >}$$

- For n representative of the whole world, this would give path lengths as large as 10-20. Even though 10 degrees of separation does not sound as impressive, it is still small.
- As a model, the Small World Model is obviously primitive as a “Systems Formation” Model. For this phenomena/purpose (explaining Milgram’s experiment), is this its most serious shortfall?

Search and Navigation

- Search
  - "To look over carefully in order to find something, to explore", "to make an effort to find something" seek, hunt, quest.
  - Network literature: **"to find the node containing information that is desired"**
- Navigate
  - "To plan, record and control the position of.." "to follow a planned Course" or "*to make one's way*"
  - Network literature: "**to get from one to another specific node by a(the) short(est) path using *only local information***"

MIT ESD
Massachusetts Institute of Technology
Engineering Systems Division

What did Milgram really show?

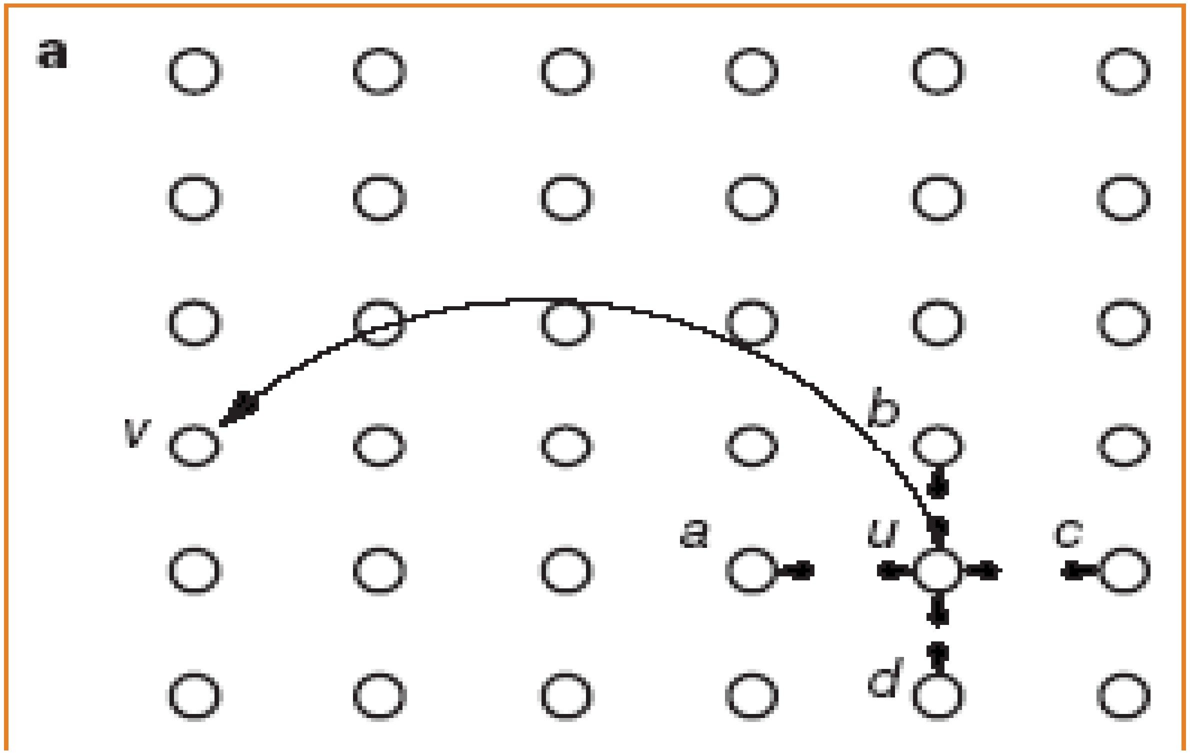
- Of 300 letters in original experiment, only 96 (random Nebraska) sampled tested the "***everyone***" part of what "***everyone***" knows
- Only 18 of these were ***ever returned*** (the preceding graph contained very "non-random" Nebraska letters)
- Other trials that were random and tested the everyone basis had even ***smaller return rates*** than Milgram's initial experiment

- Issues
  - Is ***everyone*** really 6 or less steps from ***everyone***?
  - How does ***anyone*** route such a request?
  - Apathy vs. possibility –even harder now that we all toss/delete junk mail and return rates near 1% are apparently now the norm.

Kleinberg's initial model

- Most important insight
  - Milgram's experiment did not only show that short paths exist but more importantly ***that people can*** (at least sometimes and in some circumstances) ***find them***.
- Model assumptions
  - Small World (with shortcuts added onto a **lattice** of connections) –**not randomly** but with a probability that depends on distance,

"Navigation in Small Worlds: It is easier to find short chains in some networks than others"



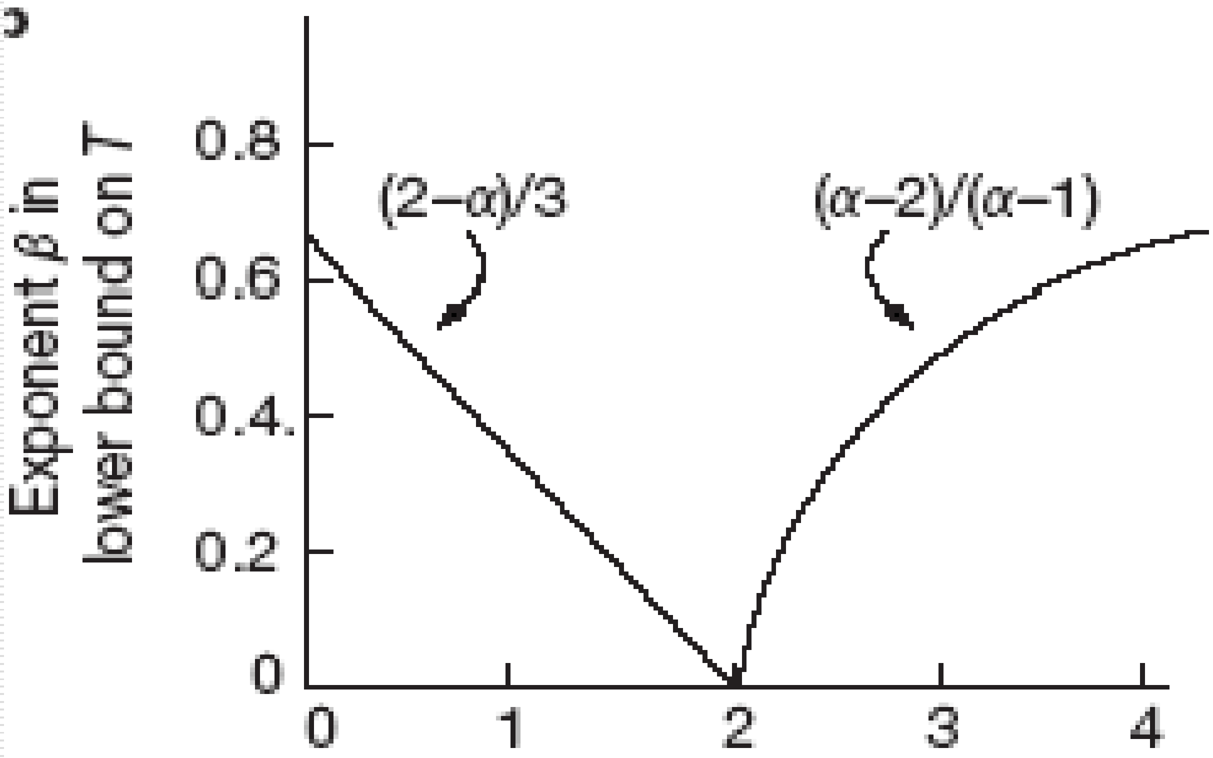
Kleinberg's initial model

- Most important insight
  - Milgram's experiment did not only show that short paths exist but more importantly ***that people can*** (at least sometimes and in some circumstances) ***find them***.
- Model assumptions
  - Small World (with shortcuts added onto a **lattice** of connections) –**not randomly** but with a probability that depends on distance, $p_s \propto r^{-\alpha}$
  - $S_{mean} \geq cn^{\beta}$ steps to find

$$\beta = (2-\alpha)/3 \ for \ \alpha < 2 \quad and$$

$$\beta = (\alpha - 2)/(\alpha - 1) \ for \ \alpha > 2$$

MIT ESD
Massachusetts Institute of Technology
Engineering Systems Division



Exponent β in lower bound on T
0.8
0.6
0.4
0.2
0
(2−α)/3
(α−2)/(α−1)
0
1
2
3
4

Results of Kleinberg I

- The existence of short paths does not guarantee that they can be found with local information
- It takes network structure of a certain kind (α = 2) to be able to do this and to get Milgram's result
- The structure Kleinberg showed worked seems quite artificial but it was a start because it showed that networks can be designed that allow for rapid search with "greedy" algorithms based on local information ("gossip" algorithms)
- Based on this work, even if ***everyone*** was connected to ***everyone***, it is surprising that ***anyone*** could find the short path.
  - Thus Milgram's famous result is ***not explained*** by this model

Next Generation structural models for navigation

- Kleinberg and independently Watts, Dodds and Newman later proposed ***a structure*** that allows such search and appears consistent with the structure of social networks.
- How would you try to route a letter to a stockbroker in Omaha?
- This structure is derived starting from clues from the "Reverse Small World Experiments" which indicate how people actually navigate social networks
  - by looking for common "features" between their targets and their acquaintances
- This structure introduces ***hierarchy into the social network*** and defines a "social distance".

Assumptions in 2ndG Navigation/Search Models I

- 1. Individuals have ***links and identities***
- ***2.*** Individuals partition the world (identities of others) into a layered hierarchy and ***distance,*** x_{ij} is assumed to be the height of the lowest common parent. The branching ratio, b, and levels, *l* define this abstraction

Assumptions in 2ndG Navigation/Search Models II

- 1. Individuals have ***links and identities***
- ***2.*** Individuals partition the world into a layered hierarchy and ***distance,*** x_{ij} is assumed to be the height of the lowest common parent. The branching ratio, b, and levels, *l* define this abstraction
- 3. Group membership signifies not only identity but also is a primary basis for determining social interaction:
- 4. Individuals hierarchically partition the world in more than one way and the model first assumes these distinctions are independent (Kleinberg shows this assumption can be relaxed with qualitatively similar results)
- 5. Individuals construct a measure of "social distance" which is the minimum over all dimensions between the nodes $p_x = c\exp[-\alpha x]$

MIT ESD
Massachusetts Institute of Technology
Engineering Systems Division

Assumptions in 2ndG Navigation/Search Models III

- 1. Individuals have ***links and identities***
- ***2.*** Individuals partition the world into a layered hierarchy and ***distance,*** x_{ij} is assumed to be the height of the lowest common parent. The branching ratio, b, and levels, *l* define this abstraction
- 3. Group membership signifies not only identity but also is a primary basis for determining social interaction:
- 4. Individuals hierarchically partition the world in more than one way and the model first assumes these distinctions are independent
- 5. Individuals construct a measure of "social distance" which is the minimum over all dimensions between the nodes

$$p_x = c\exp[-\alpha x]$$

- 6. Individuals forward messages based only on knowledge of their nearest neighbors and ***their identities***. Forward the message to someone closer to the target is the "greedy" or "gossip" algorithm used

Results I

- Successful search assumes a decent probability (.05) of finishing the chain even though the probability of terminating the search ***at each step*** is fairly high (0.25 or higher)

- Key result is that searchable networks occupy a broad range of parameter space (α , H) with almost all searchable networks having α >0 and H >1

Results II

- Successful search assumes a decent probability (.05) of finishing the chain even though the probability of terminating the search ***at each step*** is fairly high (0.25 or higher)

- Key result is that searchable networks occupy a broad range of parameter space (α, H) with almost all searchable networks having α

- Increasing group dimension beyond H = 1 yields a dramatic increase in search success (= reduction in delivery time) but "the improvement is lost as H increases further"

Probability of successful search

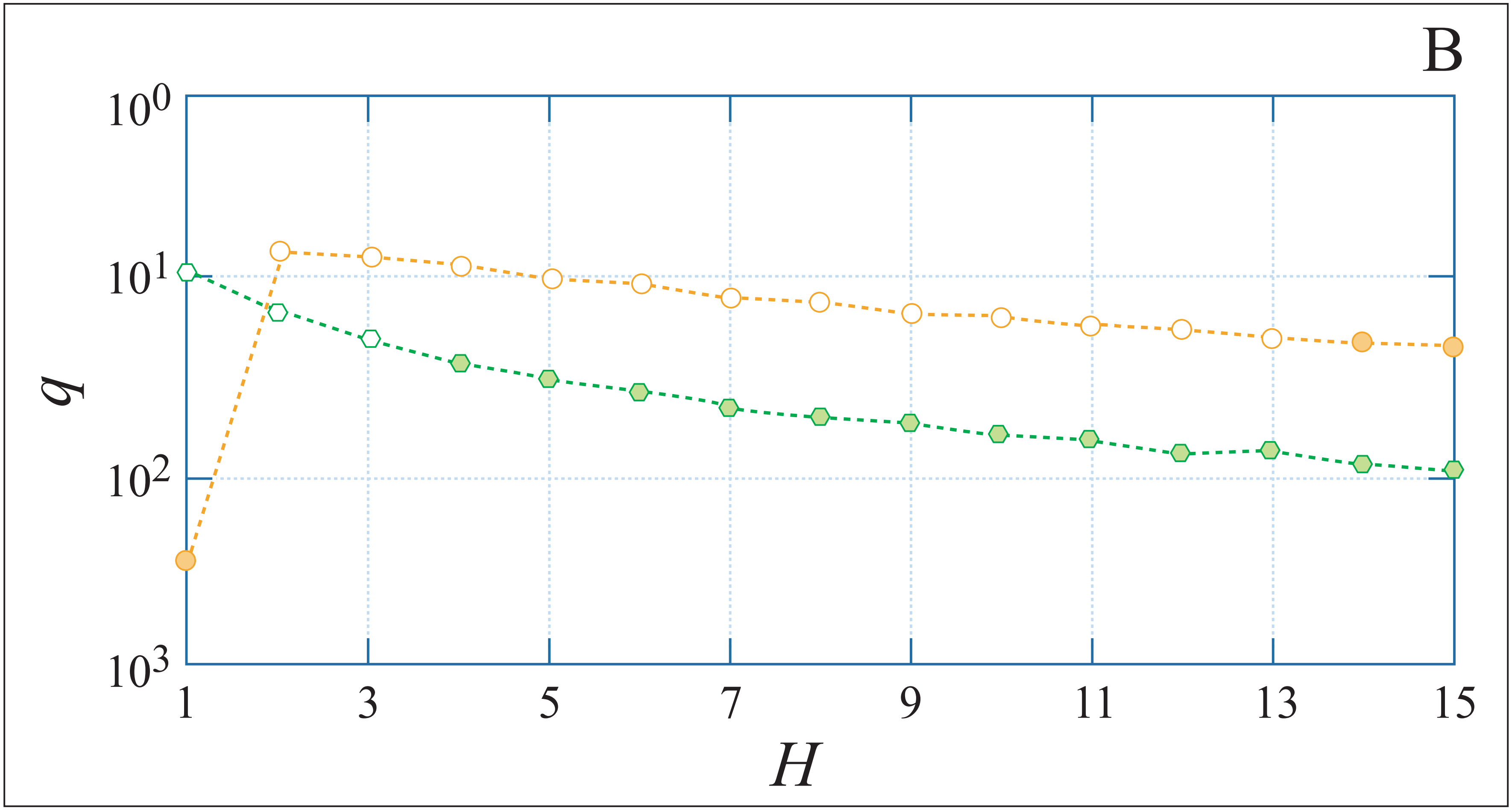


Image by MIT OpenCourseWare.

Results III

- Successful search assumes a decent probability (.05) of finishing the chain even though the probability of terminating the search ***at each step*** is fairly high (0.25 or higher)
- Key result is that searchable networks occupy a broad range of parameter space (α , H) with almost all searchable networks having α >0 and H >1
- Increasing group dimension beyond H = 1 yields a dramatic increase in search success (= reduction in delivery time) but the improvement is lost as H increases further
- For plausible values of all parameters, agreement with Milgram results are found

Distribution predicted vs. Milgram distributions

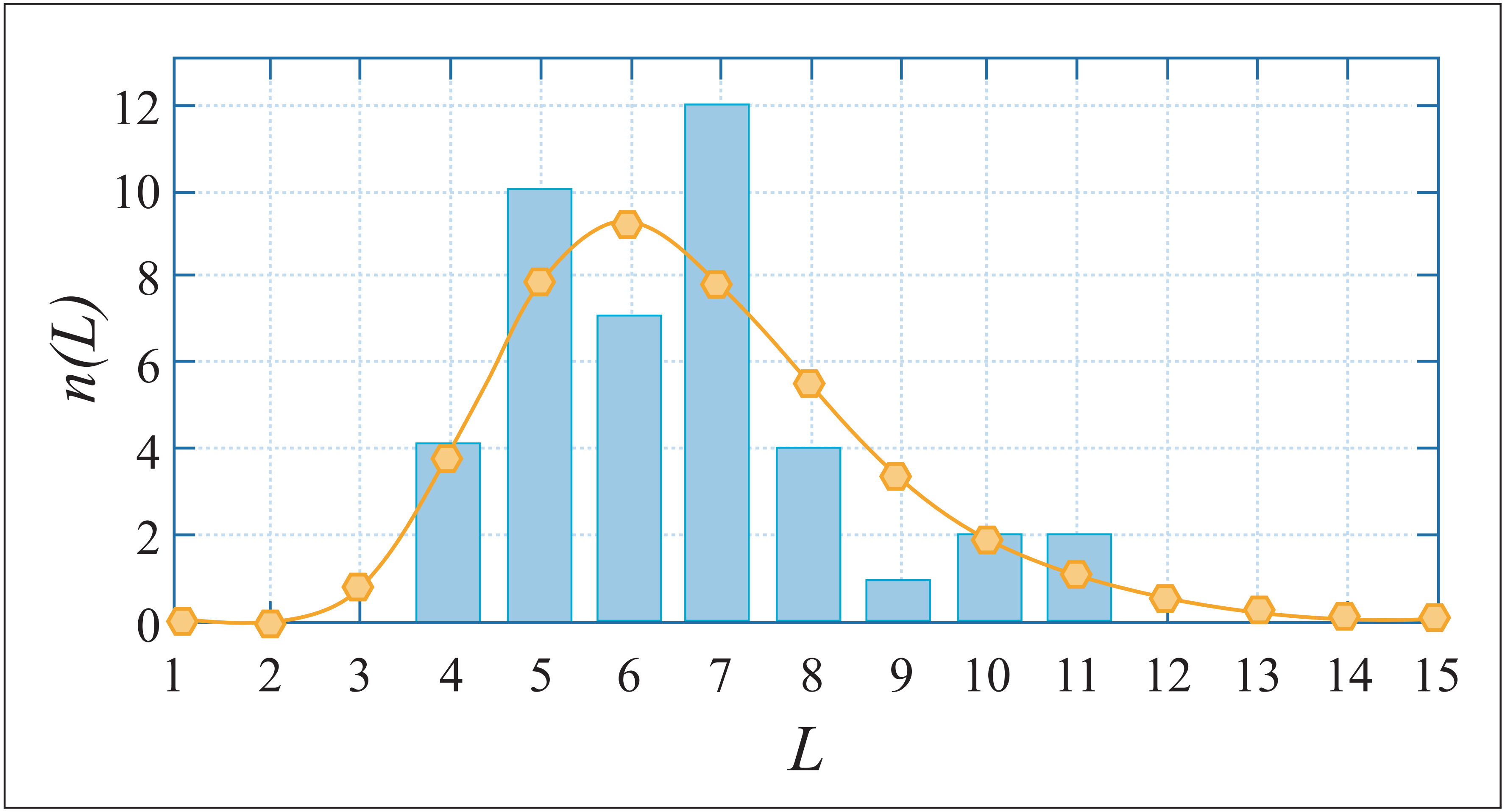


Image by MIT OpenCourseWare.

MIT ESD
Massachusetts Institute of Technology
Engineering Systems Division

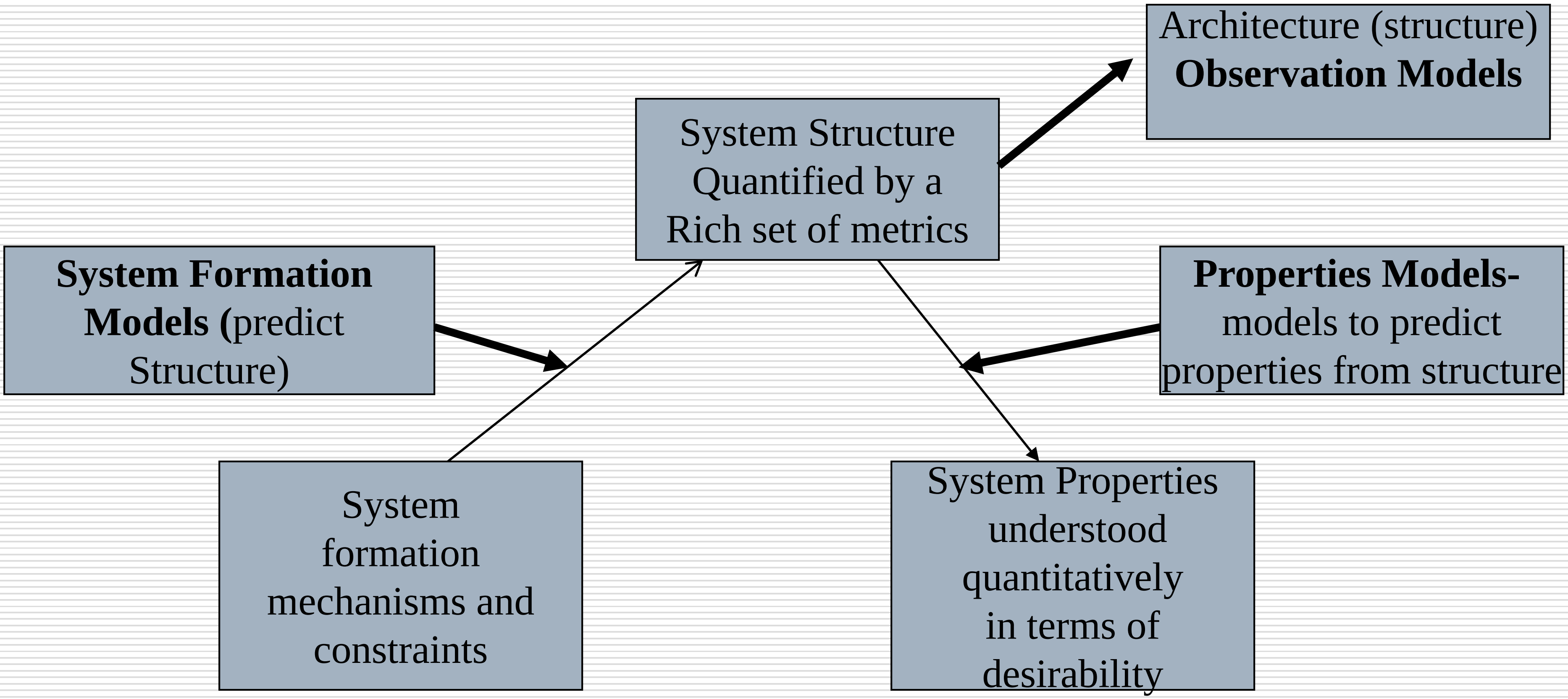
Class 14 Lecture outline

- Decomposition
  - Link to modularity
  - Practical and theoretical importance
  - Taxonomy and examples
  - Approaches to Quantitative Decomposition
    - Structural or cohesive decomposition
    - Functional decomposition
      - Roles, positions and hierarchy
      - Motifs and course graining
- → Overview of modeling

The Materials Science Metaphor

- PROCESSING> STRUCTURE> PROPERTIES
- **Structure** determines/affects **properties**
  - **Structure** is a multi-dimensional term that includes many scales and concepts simultaneously (and thus is not a “simple invisible”)
  - **Properties** include attributes that encompass dynamics, behavior and “ilities”.
  - Relationships between Structure and Properties are plentiful and became strongest as material classes under detailed study increased
  - Solid Mechanics, dislocation theory, atomic theory are some of the key enablers for deriving mechanisms to propose structure/property relationships in materials.
- In materials, properties of interest (almost always) **simultaneously** depend on **several** structural parameters. There is every reason to believe that engineering systems will similarly require numerous structural parameters to make real progress.

Schematic of Engineering System Model Types within a Framework



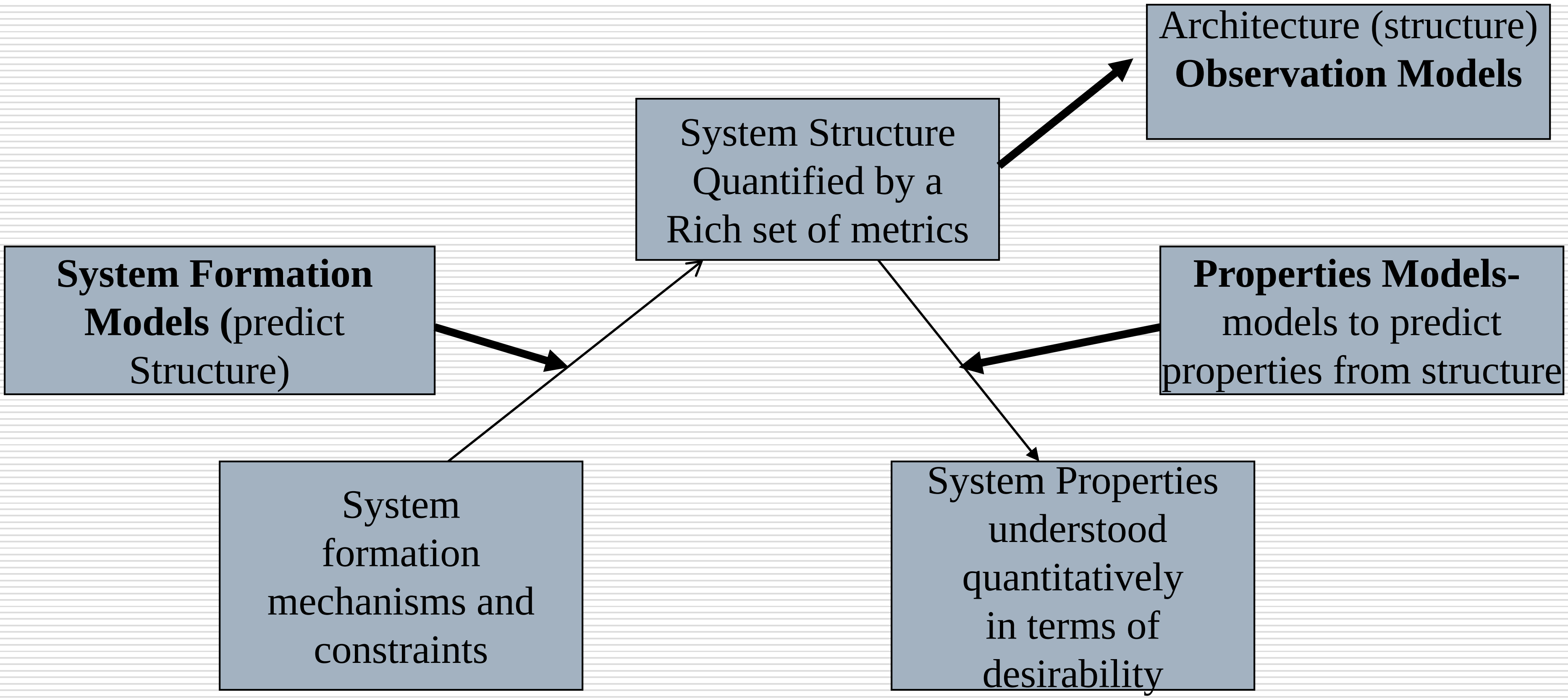
The Materials Science Metaphor II

- **Processing** determines **Structure**
  - Different Processing Modes (e-beam deposition, casting, forging, crystal growth, etc.) have different ***control parameters*** (Temperature gradient, stresses, pressure, magnetic and electrical fields, composition, etc.) that affect/determine properties.
  - ***Design*** is thus *modifying the processing modes and control parameters to obtain the desired combination of properties*. ***Understanding structure is the chief enabler of effective design***
  - Thermodynamics, phase transformations, thermal and fluid sciences, solid mechanics are useful fundamentals underlying Process/structure relationship
- Linking the framework to Engineering Systems requires discussing the structure and properties analogues in such systems.

The Materials Science Metaphor III

- Structure Characterization
  - Materials-Multiple Dimensional and very broadly construed
- Engineering Systems Possibilities for Architecture Characterization.. are also very broad (but nonetheless almost surely needs to grow)
- Engineering System Properties are also numerous (but some of the most important are not yet adequately quantified)
  - Robustness (congestion, failure of nodes and links etc.)
  - Flexibility
  - Rates of propagation (disease, ideas etc.)
  - Performance efficiency
- The **Processing > Structure > Properties** “Mantra” from materials becomes for engineering systems
  - **Formation mechanisms + constraints > architecture (structure) > Properties (ilities +)**

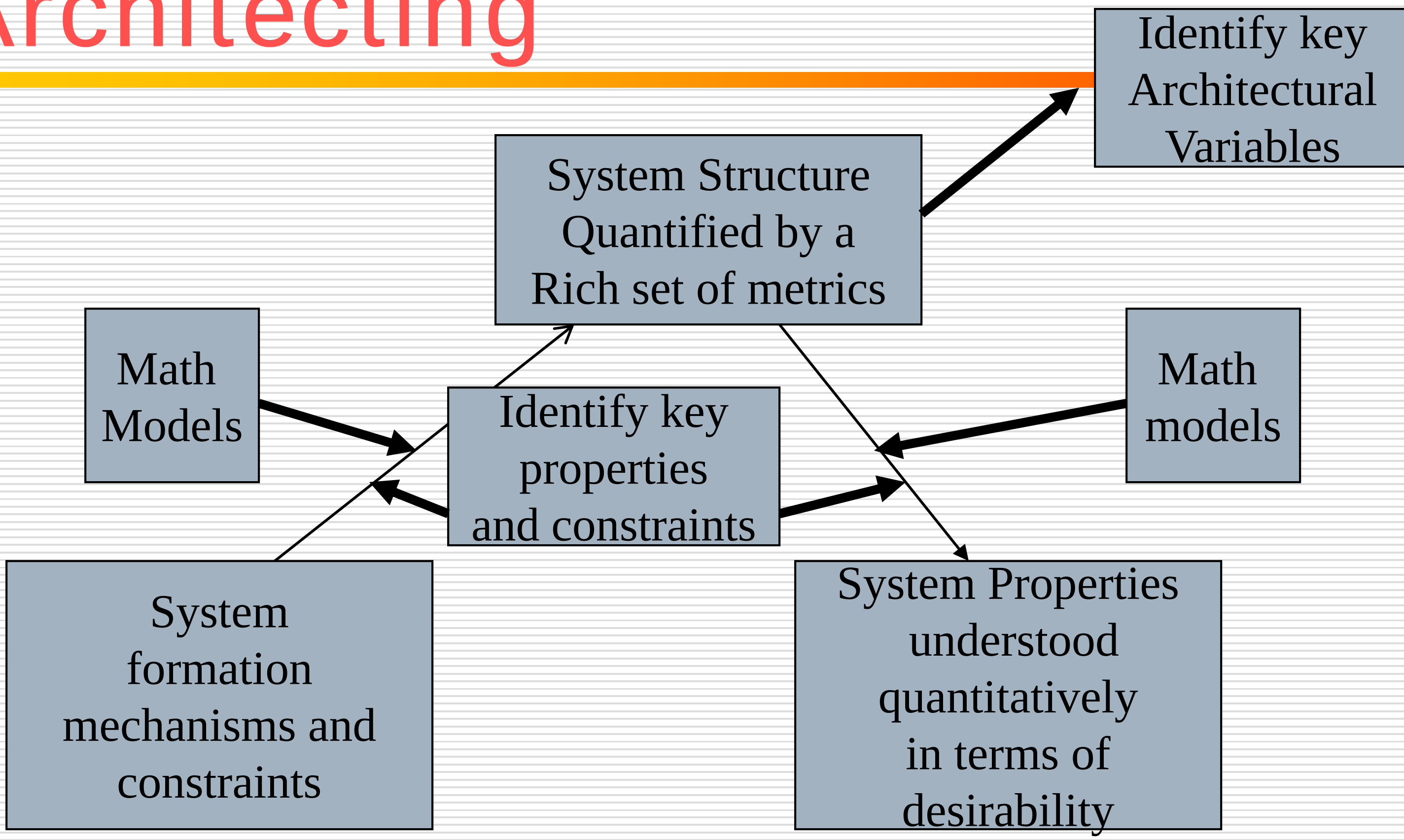
Schematic of Engineering System Model Types within a Framework



Model types

- **Models of Systems (networks)**
- **Models for predicting / explaining Structure**
  - Models for formation/growth processes of systems
  - Most network models such as random, small-world etc. implicitly fall in this category
  - Cumulative advantage, preferential attachment, bipartite community formation, heuristic optimization relative to constraints, hierarchy (or heuristics) + random
- **Models for predicting / explaining properties of systems**
  - Predicting properties from structure – architecture
    - Flexibility, robustness, performance of functions
  - Operational processes or functions
    - Communication, problem solving, decision-making, learning
    - Search and navigation
    - Failures and cascades, epidemics
- **Models / algorithms used to "observe" systems (Most Important?)**
  - Calculation of structural metrics, decomposition by cohesion and roles
  - Communities, motifs, coarse-graining, hierarchy

Schematic of Complex System Architecting

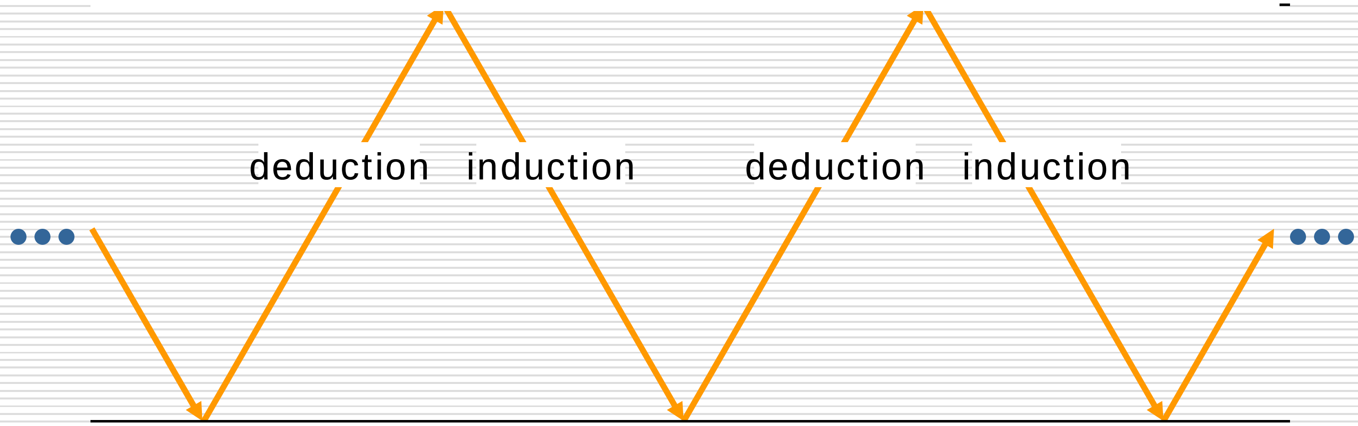


The math models of properties allow trade-off of Architectural variables and patterns of interaction on properties to drive choice of desirable structure. The math models of formation mechanisms allow choice of lowest cost or feasible sets of desirable structural metrics to be selected and evolved.

Massachusetts Institute of Technology
Engineering Systems Division

The Iterative Learning Process

Objectively obtained quantitative data (facts, phenomena)



hypothesis (model, theory that can be disproved)

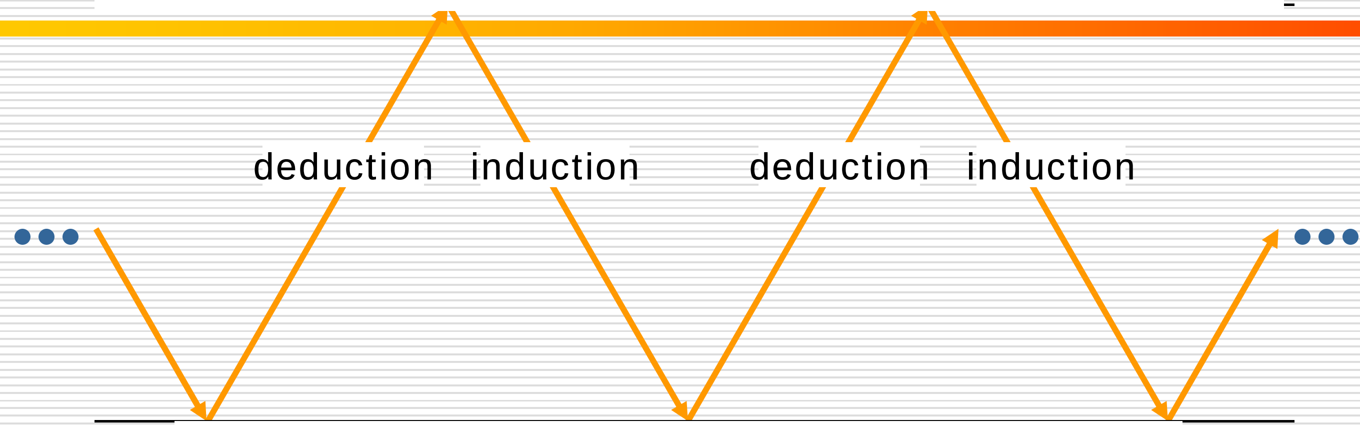
Models are “hardened” only by intensive simultaneous observational studies of relevant **reality**. The result can be

The rapid facilitation of a transition to engineering (vs. craft approaches) for the design of complex social/ technological systems

The emergence of a cumulative science in this area.

The Iterative Learning Process

Objectively obtained quantitative data (facts, phenomena)



hypothesis (model, theory that can be disproved)

Models are "hardened" only by intensive simultaneous observational studies of relevant **reality**.

••• What social distance (communication) exists in real social networks?
Random network models indicate relatively short paths might exist.
Milgram does an experiment and short paths (small worlds) exist.
Random networks do not describe clustering and short paths
Small world model is consistent and ubiquitous- Milgram experiment is revisited
Kleinberg points out navigation issue and introduces a model which treats it
but does not agree with Milgram. A 2nd generation navigation model
introduces structure into the social network and agrees with Milgram result.

Possible Future Research - Sociological Network Models

- Refinement of sociological network models
  - **Clear measurement of identity hierarchies**
  - Add **strength of ties** by methodology Newman has developed for weighted networks
- Collaborative Problem Solving in Large Organizations
  - **Community** by knowledge area vs. collaboration by problem content vs. collaboration by previous success vs. collaboration by other social network effects
- Collaboration by Internet (WWW)
  - Social Identity Hierarchy vs. non-internet
  - Interest Groups vs. age and economics
- Social Networks within organizational hierarchies
  - Identification of important characteristics that determine such networks (age, hiring group, educational institution, neighborhood, functional specialty, co-workers, etc.) and possible role/utility in organizational architecture and effectiveness
  - Influence on conflict and cooperation in organizations

Possible Future Research and Applications of Sociological Network Models b.

- Marketing Research
  - How congruent are groupings that are made in marketing research with the social network communities?
  - Can one use known communication and search results to design more effective marketing/advertising strategies?
- Stakeholder Analysis
  - Should we think of stakeholders as part of a larger sociological network?
  - What social relationships (and hierarchies) exist among different stakeholders?
- Study of sociological networks over time
  - Permanence of identity
  - Influence of communication technology on identity

References for Lecture 15

- Chapter 5 in *Six Degrees*
- J. Kleinberg, “Navigation in a Small World”, *Nature*, (2000)
- J. Kleinberg, “Small-World Phenomena and the Dynamics of Information”, *NIPS,* (2001)

- Watts,D. J., Dodds, P. S. and M. E. J. Newman, “Identity and Search in Social Networks” *Science,* 296 (May 2002).