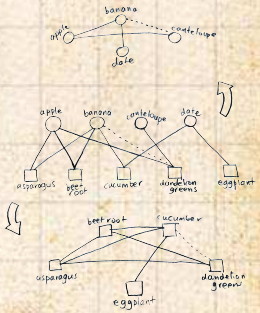
Random Bipartite Networks

Last updated: 2018/03/23, 20:59:06

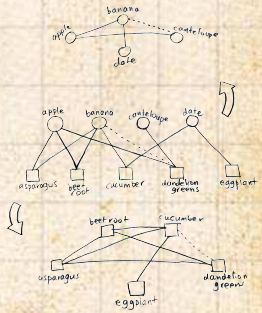
Complex Networks | @networksvox
CSYS/MATH 303, Spring, 2018

Prof. Peter Dodds | @peterdodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont



These slides are brought to you by:

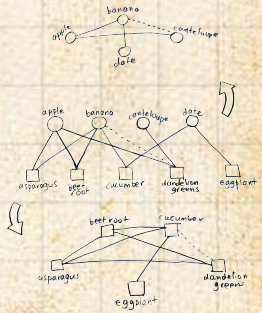


The UNIVERSITY of VERMONT

These slides are also brought to you by:

Special Guest Executive Producer

On Instagram at pratchett_the_cat



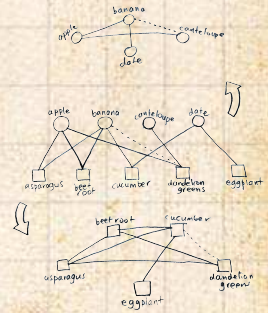
The University of Vermont

Outline

Introduction

Basic story

References




THE SUN

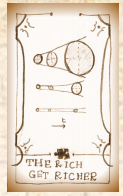
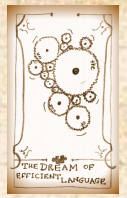
Au
THE GOLDEN AGE OF REDUCTIONISM

THE MANIFESTO

THE WEIL OF SCALES

P~S^-1
THE LAWS OF DECAYING POWER

y=x^a
THE UNEXPECTED TRANSFIGURATION

MINISTRY OF RANDOM WALKS

THE RICH GET RICHER

THE DREAM OF EFFICIENT LANGUAGE

1 2
5 4 3
5 6 7 8 9
THE LAW OF DIGITS FIRST

THE SURPRISE OF BEING ROBUST YET FRAGILE

THE MISREADING OF THE BANK OF ENO

THE MANY TAILS OF MANY TAILS

THE MEASURING

THE DATA POOR

THE POURING DATA

THE CHARMS OF THE MANY WRONG PATHS

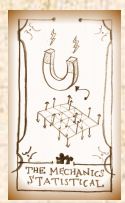
THE KNOWLEDGE OF DEAD MONKS

THE KNOWLEDGE OF DEAD MONKS

THE EMERGENCE OF STRUCTURE

THE EMERGENCE OF DESTRUCTION

THE EMERGENCE OF THINKING

It was the best of times,
THE EMERGENCE OF STORIES

THE MECHANICS OF TATISTICAL

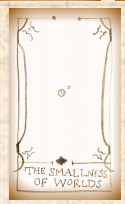
THE THRESHOLDS OF PERCOLATION

THE NETWORKS COMPLEX

THE ELEMENTS OF WEBS

THE BIRDS OF A FEATHER

THE BIRDS OF NOT A FEATHER

THE STRANGENESS OF FRIENDS

THE SMALLNESS OF WORLDS

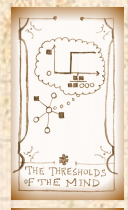
THE THEORY OF SIX DEGREES

THE NETWORKS FREE OF SCALE

CLASSIFIED
PHOENIX
BIO HAZARD
THE CONTAGION

THE CONFUSION OF CONTAGIONS

S
I
R
THE DISEASE OF RANDOM MIXING

THE UNENDING PANDEMICS

THE THRESHOLDS OF THE MIND

THE SOCIAL WILD

THE SOCIAL CONTAGION

THE RATIO OF GAIN

THE MECHANICS OF FAME

THE TALK OF FAME

THE STORIES CONTAGIOUS

THE JOHN DORY

THE ADJACENT NARRATIVES

THE STORYTELLERS

THE PATHS OF UNIVERSALITY

THE DEPENDENT PATHS

THE TROUBLES OF BIFURCATING REALITIES

?
THE UNKNOWN MECHANISM

THE LANDSCAPE OF FORKING PATHS

THE NETWORKS OF BLOOD

THE TREES OF REALITY

THE ORDERS OF STREAMS

THE LAWS OF BRANCHING

THE SCALING OF BRANCH HIERARCHIES

THE BRAID OR THE BRANCH

THE LAW OF OPTIMAL FORKS

THE CHURCH OF QUARTERLOGY

D-1
D
THE TRUTHICIDE

THE FAMILY OF PANCAKE COWS

THE FAMILY OF SPHERICAL COWS

THE IMPROBABLE PACHYDERM

THE ILL-SCALING OF ALLOMETRIC BODIES

THE SUPPLY MOST FAIR

THE SUPPLY TO THE MOST

THE DARK ART OF GENERATING FUNCTIONS

THE RANDOM SUM OF RANDOMNESS

THE TWO-SIDED WORLDS

THE STORIES AND THE TROPES

THE TRUTH OF DIFFUSION

THE UNVEILING OF HIDDEN POWERS

THE SCRYING FOR THE SHAPES OF THINGS

THE LIMITS OF ORGANIZATIONS

THE OVERCOMING OF CATASTROPHE

THE TROUBLES ILL-FORMED

THE NETWORKS OF MANY ILKS

THE CHAOS OF FETTERED CONTAGION

THE END

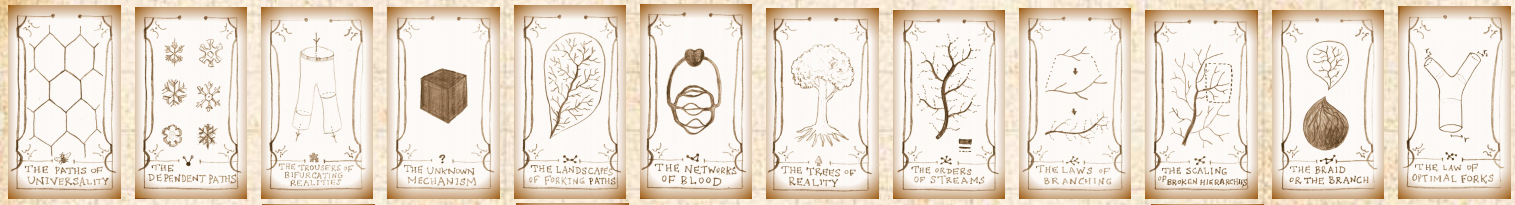
"I am the Monarch of the Sea.."
N,m
THE ARK OF THE RANDOM NETWORK


THE SUN
THE GOLDEN AGE OF REDUCTIONISM
THE MANIFESTO
THE WELL OF SCALES
THE LAWS OF DECAYING POWER
THE UNEXPECTED TRANSFIGURATION
MINISTRY OF RANDOM WALKS
THE RICH GET RICHER
THE DREAM OF EFFICIENT LANGUAGE
THE LAW OF DIGITS FIRST
THE SURPRISE OF BEING ROBUST YET FRAGILE
THE MISREADING OF THE BANK OF 2ND


THE MANY TAILS OF MANY TAILS
THE MEASURING
THE DATA POOR
THE POURING DATA
THE CHARMS OF THE MANY WRONG PATHS
THE KNOWLEDGE OF DEAD MINKS
THE KNOWLEDGE OF DEAD MINKS
THE EMERGENCE OF STRUCTURE
THE EMERGENCE OF DESTRUCTION
THE EMERGENCE OF THINKING
THE EMERGENCE OF STORIES
THE MECHANICS OF STATISTICAL

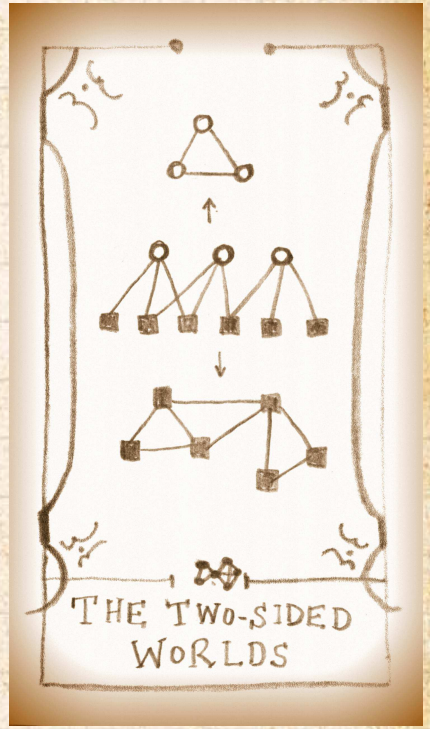

THE THRESHOLDS OF PERCOLATION
THE NETWORKS COMPLEX
THE ELEMENTS OF WEBS
THE BIRDS OF A FEATHER
THE BIRDS OF NOT A FEATHER
THE ARK OF THE RANDOM NETWORK
THE STRANGENESS OF FRIENDS
THE SMALLNESS OF WORLDS
THE THEORY OF SIX DEGREES
THE NETWORKS FREE OF SCALE
CLASSIFIED PHOENIX BIO HAZARD
THE CONTAGION
THE CONFUSION OF CONTAGIONS


THE DISEASE OF RANDOM MIXING
THE UNENDING PANDEMICS
THE THRESHOLDS OF THE MIND
THE SOCIAL WILD
THE SOCIAL CONTAGION
THE RATIO OF GAIN
THE MECHANICS OF FAME
THE TALK OF FAME
THE STORIES CONTAGIOUS
THE JOHN DORY
THE ADJACENT NARRATIVES
THE STORYTELLERS


THE PATHS OF UNIVERSALITY
THE DEPENDENT PATHS
THE TROUBLES OF BIFURCATING REALITIES
THE UNKNOWN MECHANISM
THE LANDSCAPE OF FORKING PATHS
THE NETWORKS OF BLOOD
THE TREES OF REALITY
THE ORDERS OF STREAMS
THE LAWS OF BRANCHING
THE SCALING OF BROKEN HIERARCHIES
THE BRAID OR THE BRANCH
THE LAW OF OPTIMAL FORKS


THE CHURCH OF QUARTEROLOGY
THE TRUTHICIDE
THE FAMILY OF PANCAKE COWS
THE FAMILY OF SPHERICAL COWS
THE IMPROBABLE PACHYDERM
THE ILL-SCALING OF ALLOMETRIC BODIES
THE SUPPLY MOST FAIR
THE SUPPLY TO THE MOST
THE DARK ART OF GENERATING FUNCTIONS
THE RANDOM SUM OF RANDOMNESS


THE STORIES AND THE TROPES
THE TRUTH OF DIFFUSION
THE UNVEILING OF HIDDEN POWERS
THE SCRYING OF THE SHAPES OF THINGS
THE LIMITS OF ORGANIZATIONS
THE OVERCOMING OF CATASTROPHE
THE TROUBLES ILL-FORMED
THE NETWORKS OF MANY ILKS
THE CHAOS OF FETTERED CONTAGION
THE END


THE TWO-SIDED WORLDS


SCIENTIFIC REPORTS

Flavor network and the principles of food pairing

"Flavor network and the principles of food pairing"

Ahn et al.,
Nature Scientific Reports, **1**, 196, 2011. [1]

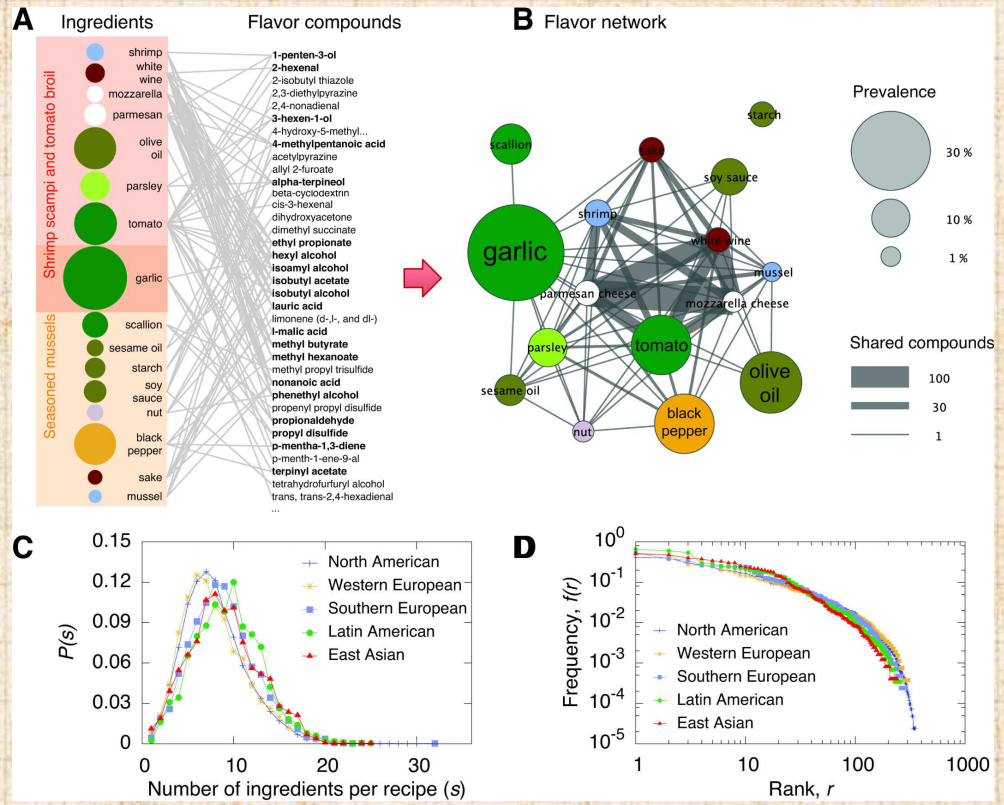
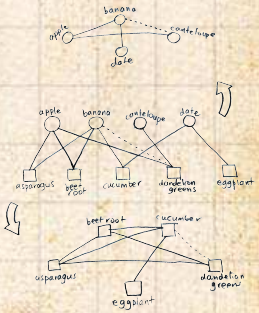


Figure 1 | Flavor network. (A) The ingredients contained in two recipes (left column), together with the flavor compounds that are known to be present in the ingredients (right column). Each flavor compound is linked to the ingredients that contain it, forming a bipartite network. Some compounds (shown in boldface) are shared by multiple ingredients. (B) If we project the ingredient-compound bipartite network into the ingredient space, we obtain the *flavor network*, whose nodes are ingredients, linked if they share at least one flavor compound. The thickness of links represents the number of flavor compounds two ingredients share and the size of each circle corresponds to the prevalence of the ingredients in recipes. (C) The distribution of recipe size, capturing the number of ingredients per recipe, across the five cuisines explored in our study. (D) The frequency-rank plot of ingredients across the five cuisines show an approximately invariant distribution across cuisines.



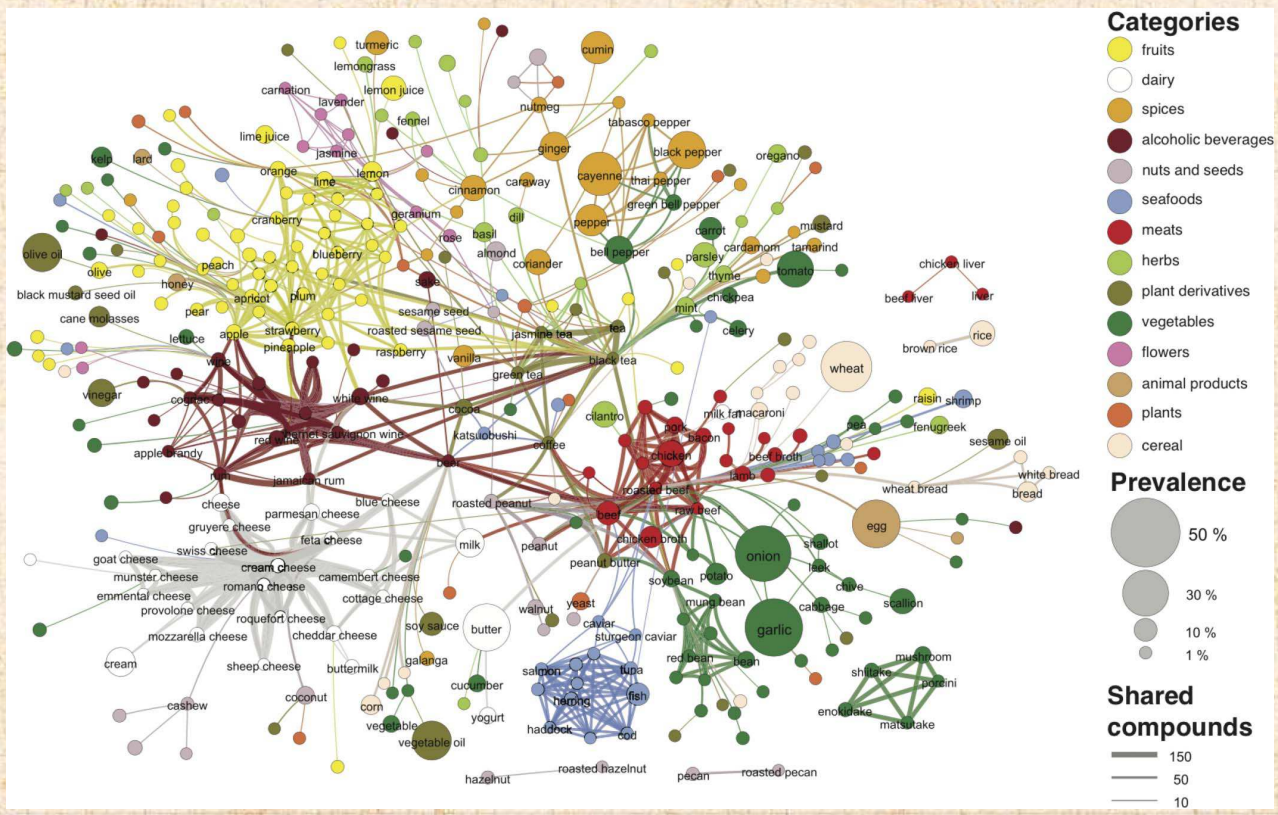
The UNIVERSITY of VERMONT


SCIENTIFIC REPORTS

Flavor network and the principles of food pairing

"Flavor network and the principles of food pairing"

Ahn et al.,
Nature Scientific Reports, **1**, 196, 2011. [1]



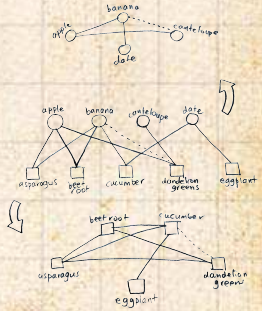
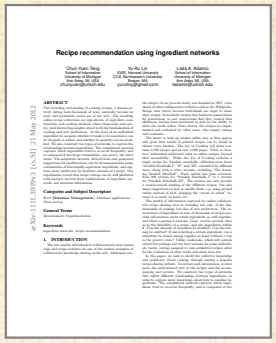


Figure 2 | The backbone of the flavor network. Each node denotes an ingredient, the node color indicates food category, and node size reflects the ingredient prevalence in recipes. Two ingredients are connected if they share a significant number of flavor compounds, link thickness representing the number of shared compounds between the two ingredients. Adjacent links are bundled to reduce the clutter. Note that the map shows only the statistically significant links, as identified by the algorithm of Refs.[28,29] for *p*-value 0.04. A drawing of the full network is too dense to be informative. We use, however, the full network in our subsequent measurements.


Recipe recommendation using ingredient networks

Chun-Yuen Teng
School of Information
University of Michigan
Ann Arbor, MI, USA
chunyuen@umich.edu

Yu-Ru Lin
College of Computer and Information Science
CCIS, Northeastern University
Boston, MA
yurulin@gmail.com

Lada A. Adamic
School of Information
University of Michigan
Ann Arbor, MI, USA
ladamic@umich.edu

ABSTRACT

The recording and sharing of cooking recipes, a human activity dating back thousands of years, is one of the most widely and passionately carried out of the arts. The resulting recipe collections have been studied from a number of different literature and cooking methods, where large-scale and computational approaches to cooking have been applied to cooking and taste preferences. At the level of an individual ingredient we measure whether it tends to be essential or optional for a recipe, and how substitutable it is with other ingredients. We also construct two types of networks to capture the flavor profiles of ingredients. The first is a substitute network, which organizes foods based on overall frequently used substitutes, revealing clusters of functionally similar ingredients. The substitute network, derived from user generated suggestions for modifications, can be decomposed into meaningful communities. The second is a complement network, which captures which ingredients tend to occur together more than would be expected by chance. The complement network, derived from statistical regularities in recipe composition, can be used to automatically suggest ingredients to complete a recipe, and allows us to quantify the extent to which different cuisines tend to use ingredients in characteristic ways. Applying machine learning techniques, we show that ingredient substitutes can be accurately predicted using nutritional profiles and complementarity relationships.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database applications—Data mining

General Terms

Measurement, Experimentation

Keywords

Association networks, recipe recommendation.

1. INTRODUCTION

The web enables individuals to collaboratively share knowledge and recipe websites are one of the earliest examples of collaborative knowledge-sharing on the web. Although recipes

the subject of our present study, was founded in 1997, years ahead of other collaborative websites such as the Wikipedia. Like the Wikipedia, Epicurious relies on its users to provide fresh recipes, functionally recipes that facilitate general desire to cook, and to share and rate recipes that they have found to be effortless. Having been mentioned in part by the ability to search recipes by ingredients, the Epicurious recipes are also reused and enhanced by other users, who simply indicate their preferences.

The choice to look up recipe online may at first appear surprising, given that we usually cook in the kitchen, with physical cook's kitchen. The day of Cooking 2.0 about who runs 4,500 recipes spread over 1,900 pages. There is, however, a practical reason for the choice to look up recipes: they are accessible. While the day of Cooking contains a number of recipes, the majority of recipes are scattered in "Standard Australia 1", "87", and "88", submitted by different users. Epicurious recipes are also "Gourmet Today" and "Bon Appétit Magazine". Each recipe has been reviewed by 500–600 readers for "Overall, Standard 1" to 5 scores. Epicurious also provides a "Recipe Box" that allows users to create a customized reading of the different recipes, but also to create a personal collection of recipes. Epicurious is a rich variety instead of both dropping the "cream of wheat" for "cream of wheat" for "cream of wheat".

The wealth of information captured by online collective recipe sharing sites is revealing and rich of this has been studied from a number of different perspectives. The observation of dependencies between ingredients is at the heart of food pairing [1], which suggests that ingredients that are shared when a pairing is unusual. Users' reviews provide clues about which recipes are likely to be successful [2] and which are not [3]. Cuisines are also studied [4]. Cuisines can be characterized by the ingredients that are over-represented [5], which is especially useful in the case of geographically distinct cuisines [6]. Finally, the study of cooking techniques [7] and the role of ingredients in the success of recipes [8] have been studied.

In this paper, we seek to distill the collective knowledge of recipe websites into a computational model of ingredients and cooking methods. To extract such information, we first collect a large number of recipes from Epicurious.com for grouping and reviews. We construct two types of networks that reflect different relationships between ingredients. In the substitute network, ingredients are connected if they are found to be used frequently as substitutes of one

"Recipe recommendation using ingredient networks"

Teng, Lin, and Adamic, Proceedings of the 3rd Annual ACM Web Science Conference, **1**, 298–307, 2012. [8]

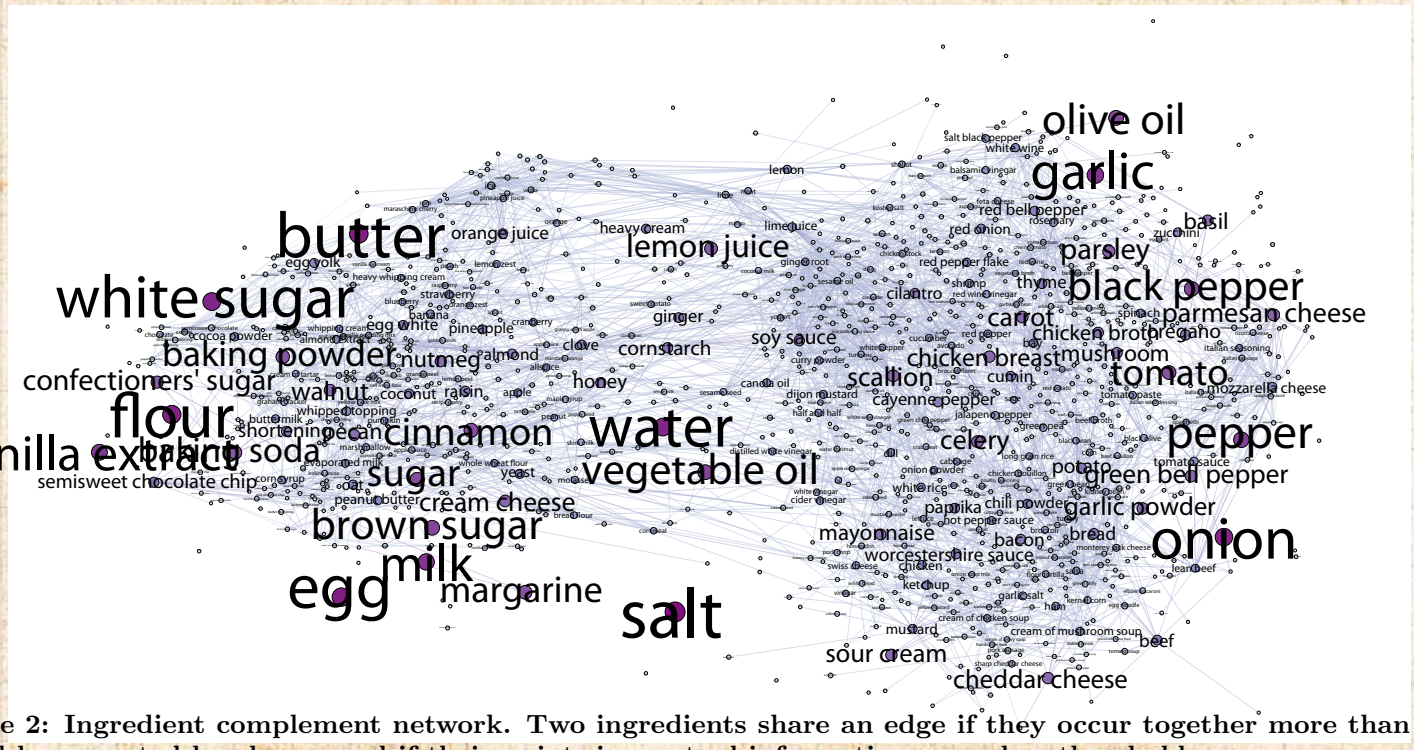
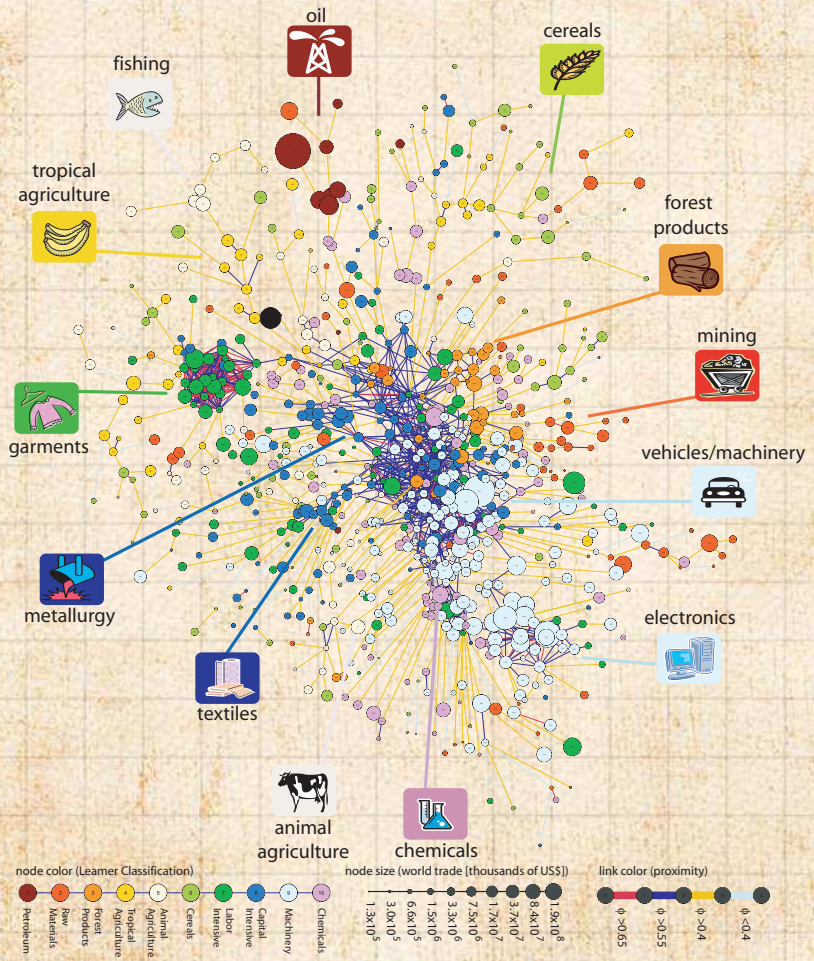
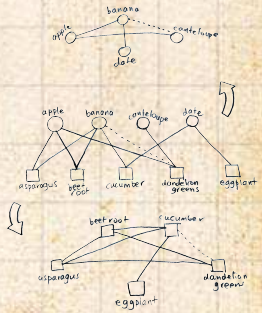


Figure 2: Ingredient complement network. Two ingredients share an edge if they occur together more than would be expected by chance and if their pointwise mutual information exceeds a threshold.

“The Product Space Conditions the Development of Nations”

Hidalgo et al.,
Science, **317**, 482–487, 2007. [6]





Networks and creativity:

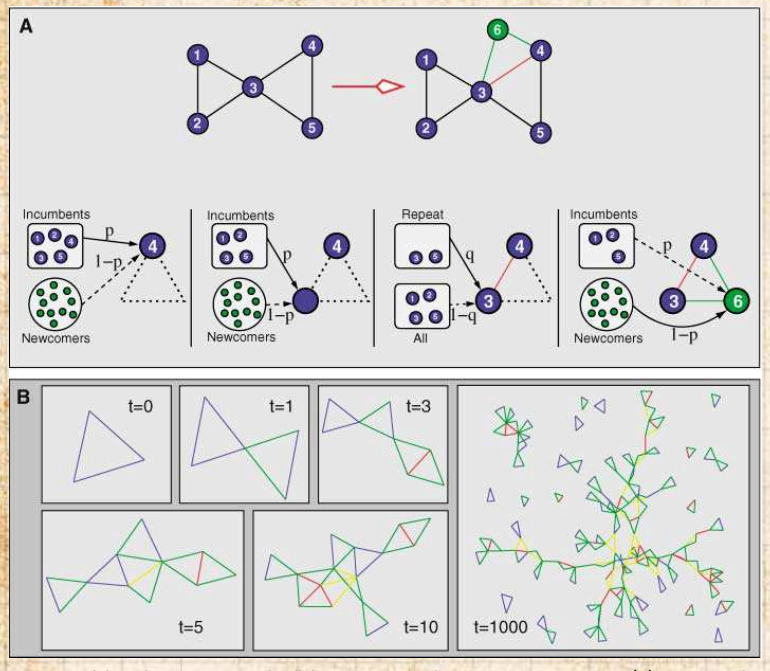
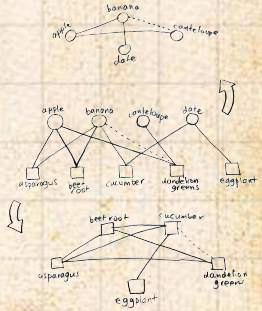


Fig. 2. Modeling the emergence of collaboration networks in creative enterprises. (**A**) Creation of a team with $m = 3$ agents. Consider, at time zero, a collaboration network comprising five agents, all incumbents (blue circles). Along with the incumbents, there is a large pool of newcomers (green circles) available to participate in new teams. Each agent in a team has a probability p of being drawn from the pool of incumbents and a probability $1 - p$ of being drawn from the pool of newcomers. For the second and subsequent agents selected from the incumbents' pool: (i) with probability q, the new agent is randomly selected from among the set of collaborators of a randomly selected incumbent already in the team; (ii) otherwise, he or she is selected at random among all incumbents in the network. For concreteness, let us assume that incumbent 4 is selected as the first agent in the new team (leftmost box). Let us also assume that the second agent is an incumbent, too (center-left box). In this example, the second agent is a past collaborator of agent 4, specifically agent 3 (center-right box). Lastly, the third agent is selected from the pool of newcomers; this agent becomes incumbent 6 (rightmost box). In these boxes and in the following panels and figures, blue lines indicate newcomer-newcomer collaborations, green lines indicate newcomer-incumbent collaborations, yellow lines indicate new incumbent-incumbent collaborations, and red lines indicate repeat collaborations. (**B**) Time evolution of the network of collaborations according to the model for $p = 0.5$, $q = 0.5$, and $m = 3$.

- Guimerà et al., Science 2005: [5] "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance"
- Broadway musical industry
- Scientific collaboration in Social Psychology, Economics, Ecology, and Astronomy.

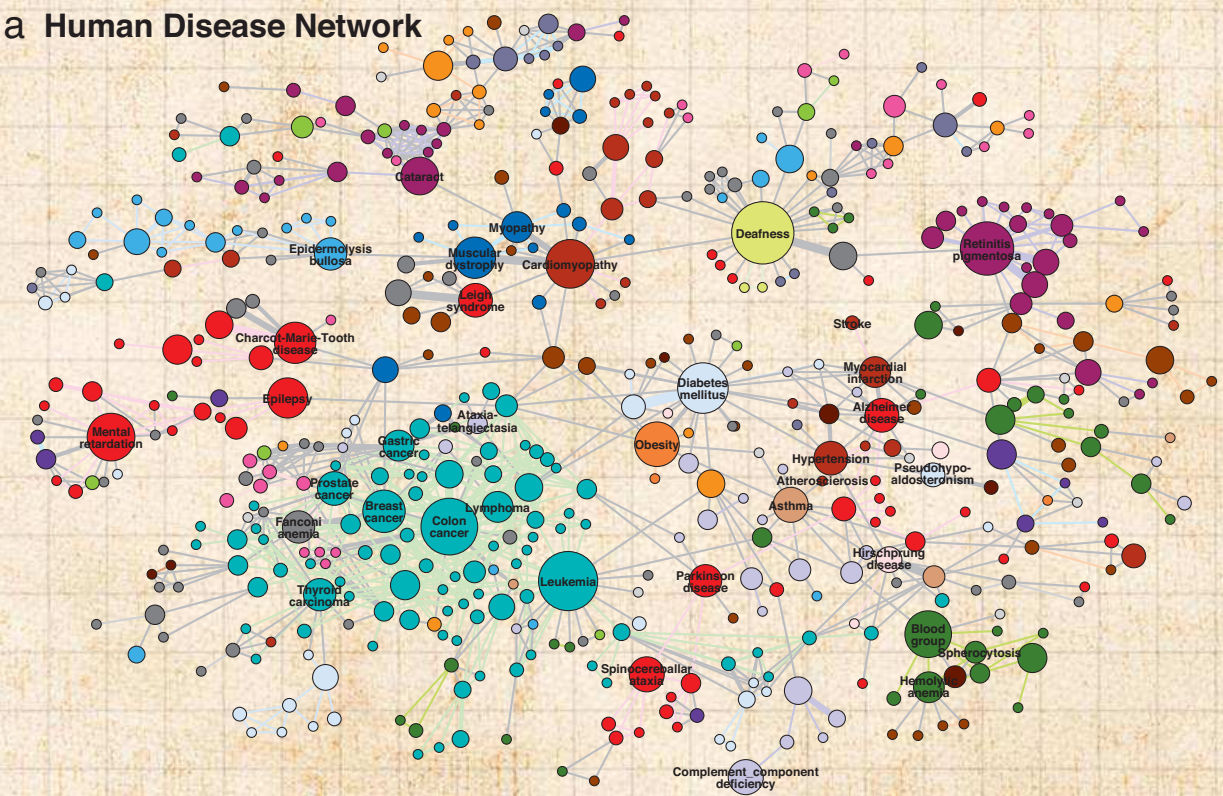


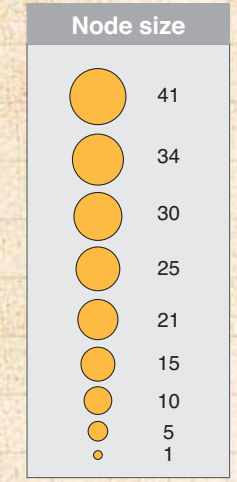

The human disease network

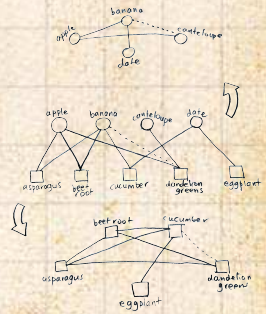
"The human disease network"

Goh et al.,
Proc. Natl. Acad. Sci., **104**, 8685–8690, 2007. [4]

a **Human Disease Network**



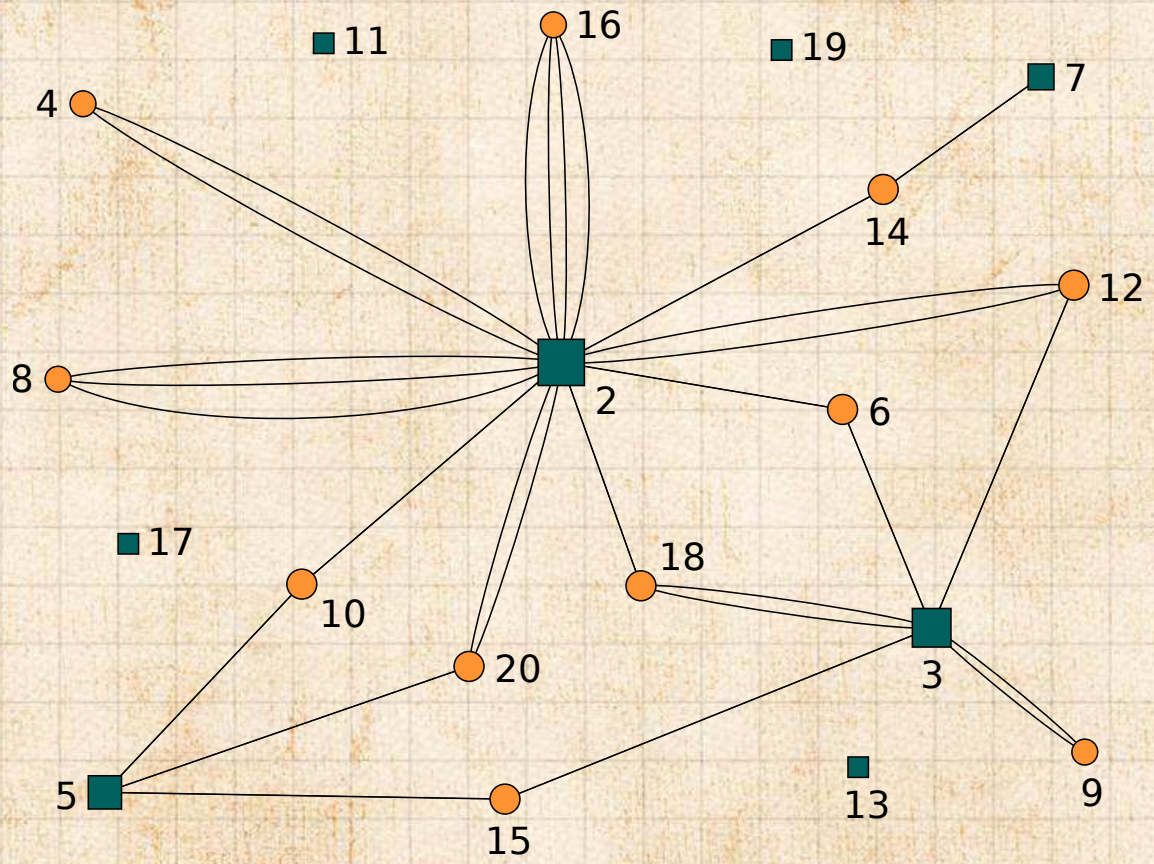


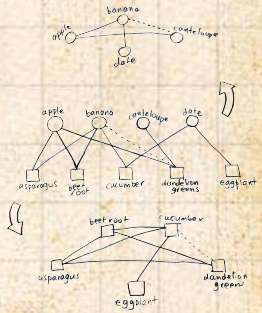


The UNIVERSITY of VERMONT

“The complex architecture of primes and natural numbers”

García-Pérez, Serrano, and Boguñá, http://arxiv.org/abs/1402.3612, 2014. [3]



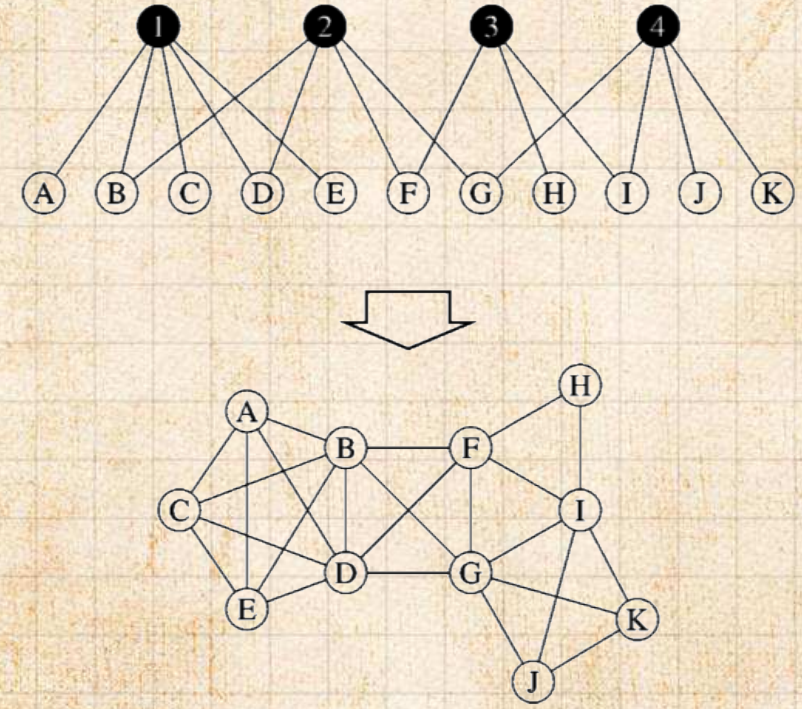


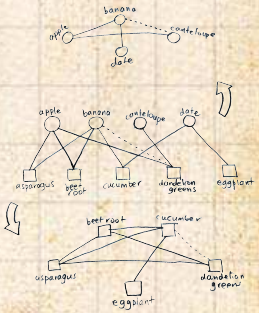
Random bipartite networks:

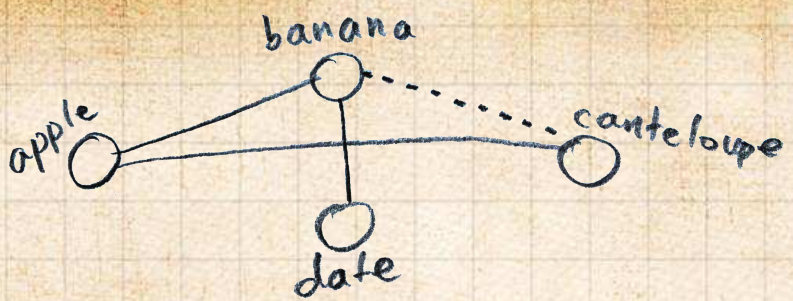
We'll follow this rather well cited paper:

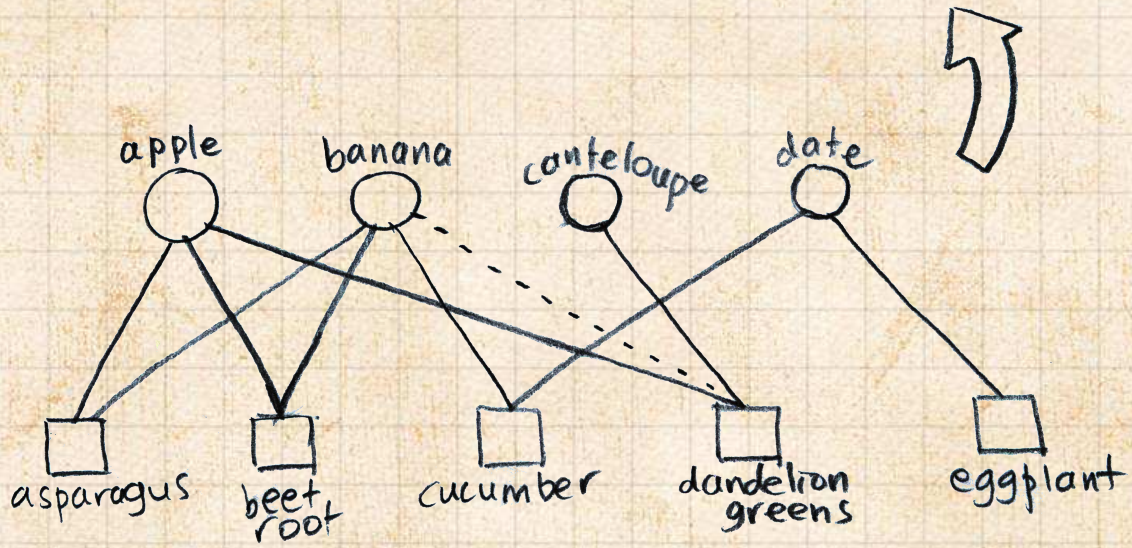
"Random graphs with arbitrary degree distributions and their applications"

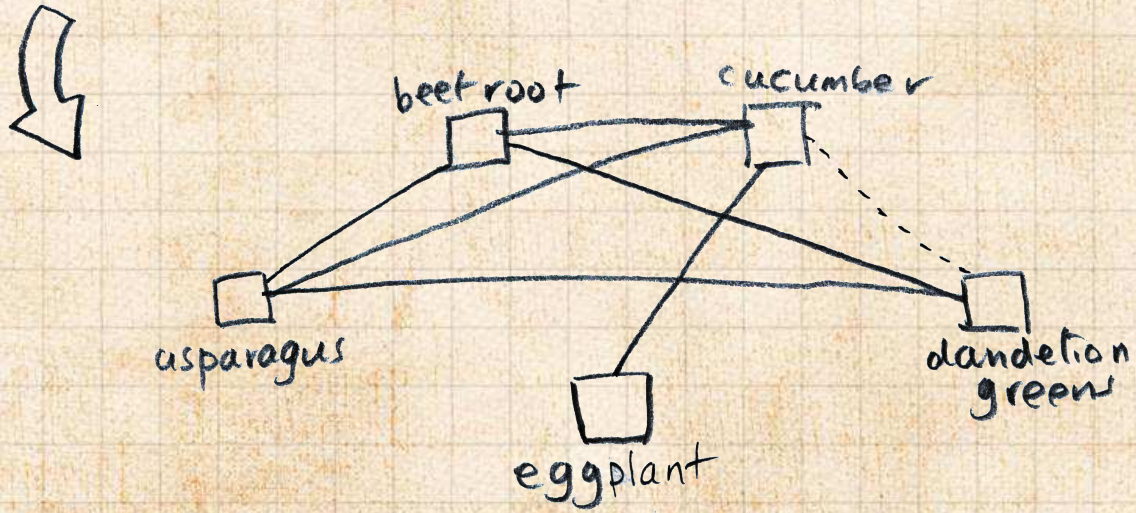
Newman, Strogatz, and Watts,
Phys. Rev. E, **64**, 026118, 2001. [7]

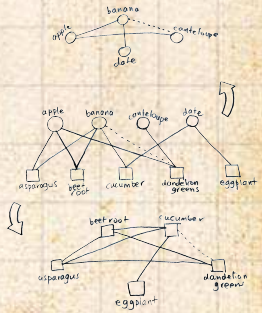




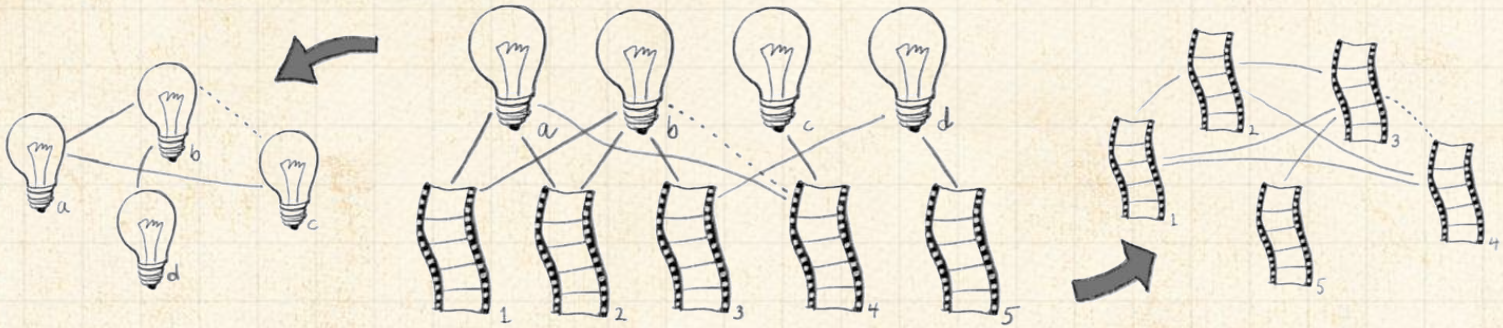




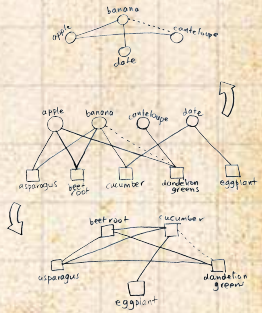




Example of a bipartite affiliation network and the induced networks:



- Center: A small story-trope bipartite graph. [2]
- Induced trope network and the induced story network are on the left and right.
- The dashed edge in the bipartite affiliation network indicates an edge added to the system, resulting in the dashed edges being added to the two induced networks.




THE SUN

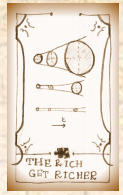
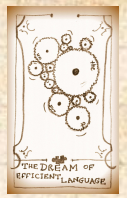
Au
THE GOLDEN AGE OF REDUCTIONISM

THE MANIFESTO

THE WEIL OF SCALES

THE LAWS OF DECAYING POWER

THE UNEXPECTED TRANSFIGURATION

MINISTRY OF RANDOM WALKS

THE RICH GET RICHER

THE DREAM OF EFFICIENT LANGUAGE

1 2
5 4 3
6 7 8 9
THE LAW OF DIGITS FIRST

THE SURPRISE OF BEING ROBUST YET FRAGILE

THE MISREADING OF THE BANK OF ENID

THE MANY TAILS OF MANY TAILS

THE MEASURING

THE DATA POOR

THE POURING DATA

THE CHARMS OF THE MANY WRONG PATHS

THE KNOWLEDGE OF DEAD MINKS

THE KNOWLEDGE OF DEAD MINKS

THE EMERGENCE OF STRUCTURE

THE EMERGENCE OF DESTRUCTION

THE EMERGENCE OF THINKING

It was the best of times,
THE EMERGENCE OF STORIES

THE MECHANICS OF STATISTICAL

THE THRESHOLDS OF PERCOLATION

THE NETWORKS COMPLEX

THE ELEMENTS OF WEBS

THE BIRDS OF A FEATHER

THE BIRDS OF NOT A FEATHER

THE ARK OF THE RANDOM NETWORK

THE STRANGENESS OF FRIENDS

THE SMALLNESS OF WORLDS

THE THEORY OF SIX DEGREES

THE NETWORKS FREE OF SCALE

CLASSIFIED
PHOENIX
BIO HAZARD
THE CONTAGION

THE CONFUSION OF CONTAGIONS

THE DISEASE OF RANDOM MIXING

THE UNENDING PANDEMICS

THE THRESHOLDS OF THE MIND

THE SOCIAL WILD

THE SOCIAL CONTAGION

THE RATIO OF GAIN

THE MECHANICS OF FAME

THE TALK OF FAME

THE STORIES CONTAGIOUS

THE JOHN DORY

THE ADJACENT NARRATIVES

THE STORYTELLERS

THE PATHS OF UNIVERSALITY

THE DEPENDENT PATHS

THE TROUSERS OF BIFURCATING REALITIES

?
THE UNKNOWN MECHANISM

THE LANDSCAPE OF FORKING PATHS

THE NETWORK OF BLOOD

THE TREES OF REALITY

THE ORDERS OF STREAMS

THE LAWS OF BRANCHING

THE SCALING OF BROKEN HIERARCHIES

THE BRAID OR THE BRANCH

THE LAW OF OPTIMAL FORKS

THE CHURCH OF QUARTEROLOGY

D=1
D
THE TRUTHICIDE

THE FAMILY OF PANCAKE COWS

THE FAMILY OF SPHERICAL COWS

THE IMPROBABLE PACHYDERM

THE ILL-SCALING OF ALLOMETRIC BODIES

THE SUPPLY MOST FAIR

THE SUPPLY TO THE MOST

THE DARK ART OF GENERATING FUNCTIONS

THE RANDOM SUM OF RANDOMNESS

THE TWO-SIDED WORLDS

THE TRUTH OF DIFFUSION

THE UNVEILING OF HIDDEN POWERS

THE SCRYING OF SHAPES OF THINGS

THE LIMITS OF ORGANIZATIONS

THE OVERCOMING OF CATASTROPHE

THE TROUBLES ILL-FORMED

THE NETWORKS OF MANY ILLS

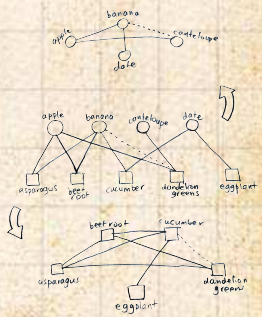
THE CHAOS OF FETTERED CONTAGION

THE
7
END

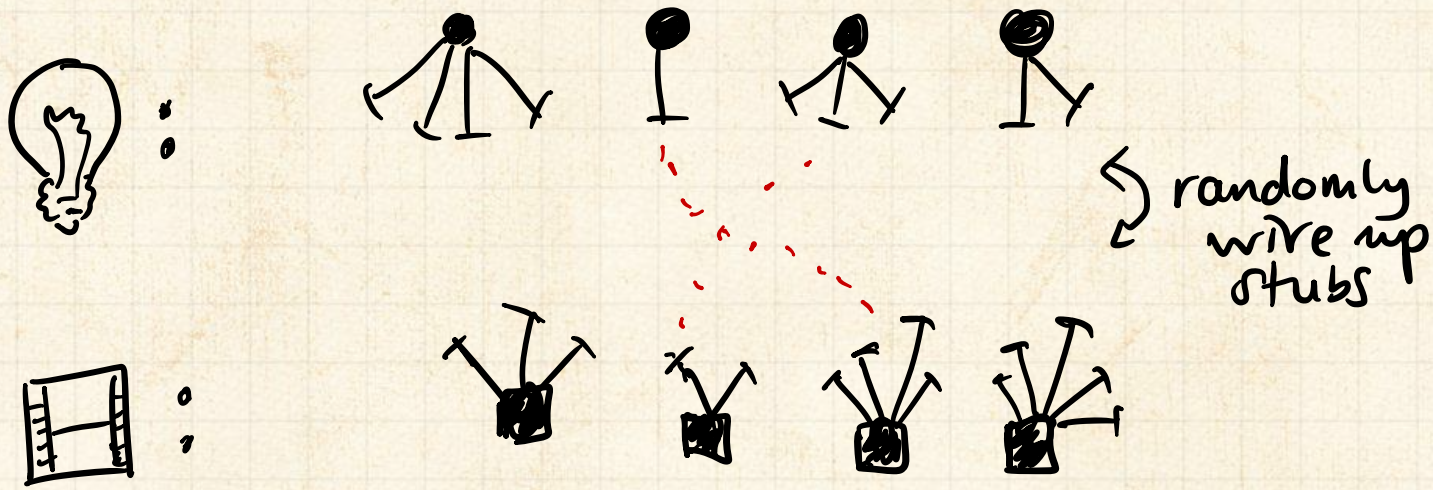
THE STORIES AND THE TROPES

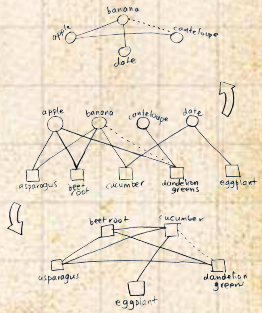
Basic story:

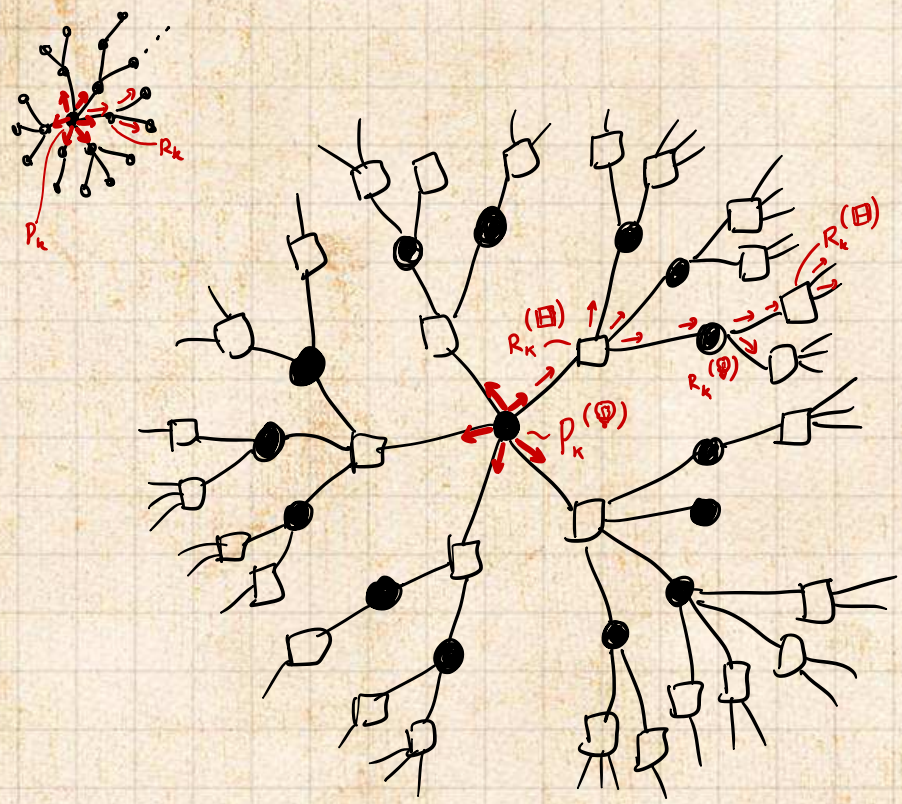
- An example of two inter-affiliated types:
  - 🎞 = stories,
  - 💡 = tropes.
- Stories contain tropes, tropes are in stories.
- Consider a story-trope system with $N_{🎞}$ = # stories and $N_{💡}$ = # tropes.
- $m_{🎞,💡}$ = number of edges between 🎞 and 💡.
- Let's have some underlying distributions for numbers of affiliations: $P_k^{(🎞)}$ (a story has k tropes) and $P_k^{(💡)}$ (a trope is in k stories).
- Average number of affiliations: $\langle k \rangle_{🎞}$ and $\langle k \rangle_{💡}$.
  - $\langle k \rangle_{🎞}$ = average number of tropes per story.
  - $\langle k \rangle_{💡}$ = average number of stories containing a given trope.
- Must have balance: $N_{🎞} \cdot \langle k \rangle_{🎞} = m_{🎞,💡} = N_{💡} \cdot \langle k \rangle_{💡}$.



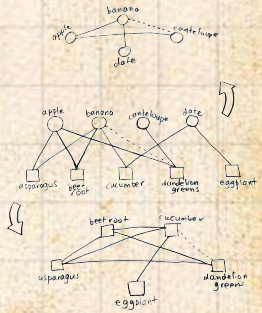
How to build:


randomly wire up stubs


banana
apple
cantaloupe
date
apple
banana
cantaloupe
date
asparagus
beet root
cucumber
dandelion greens
eggplant
beet root
cucumber
asparagus
dandelion greens
eggplant



See
Bipartite
random networks
as
Generalized
random networks
with
alternating
degree
distributions



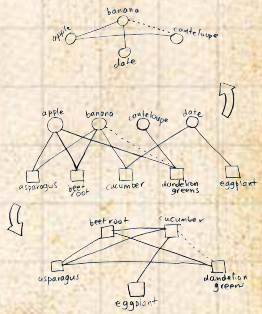
Usual helpers for understanding network's structure:

- Randomly select an edge connecting a 🎞 to a 💡.
- Probability the 🎞 contains k other tropes:

$$R_k^{(🎞)} = \frac{(k+1)P_{k+1}^{(🎞)}}{\sum_{j=0}^{N_🎞}(j+1)P_{j+1}^{(🎞)}} = \frac{(k+1)P_{k+1}^{(🎞)}}{\langle k\rangle_🎞}.$$

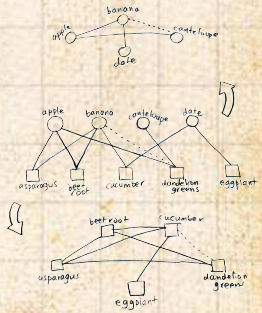
- Probability the 💡 is in k other stories:

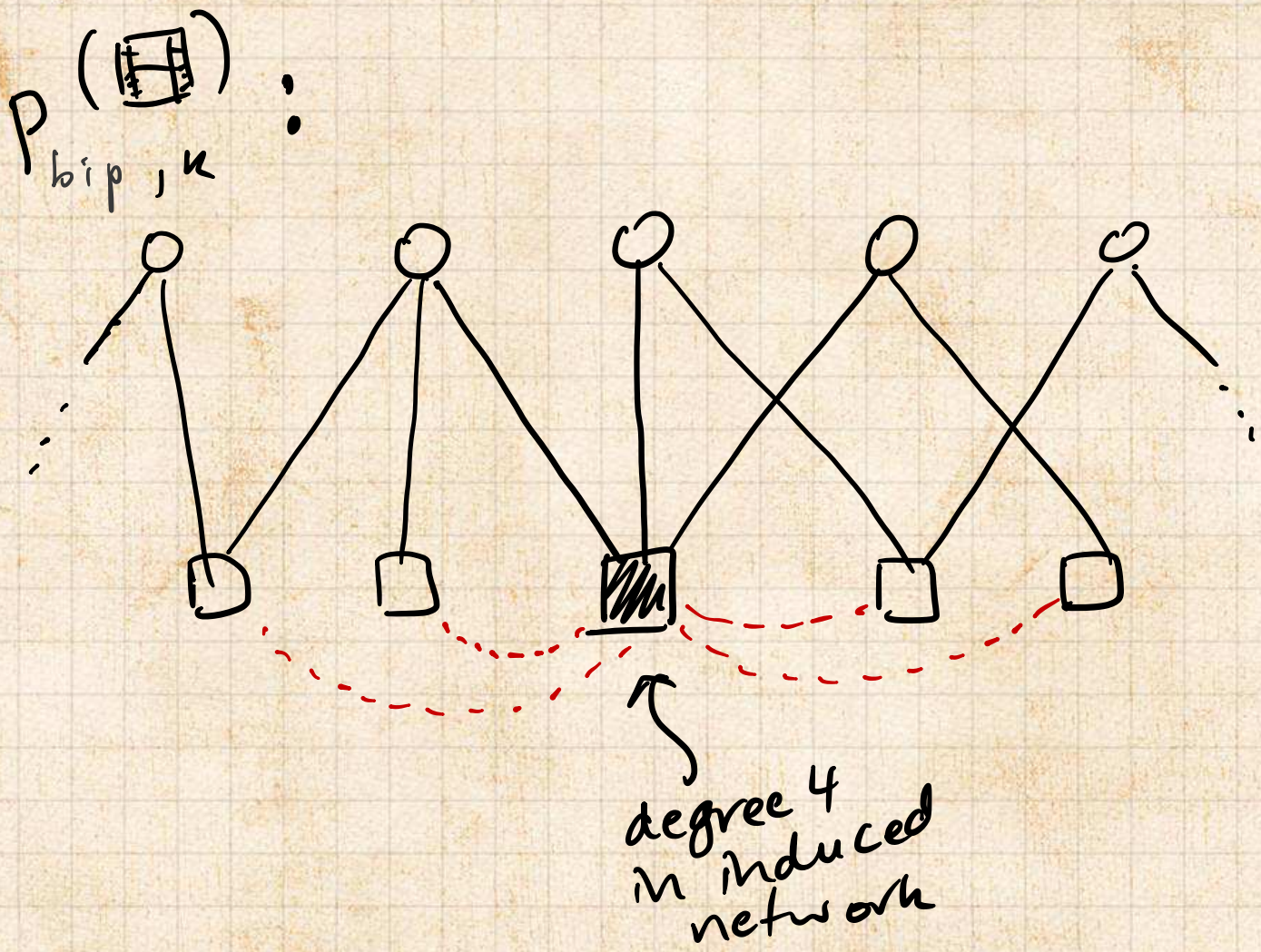
$$R_k^{(💡)} = \frac{(k+1)P_{k+1}^{(💡)}}{\sum_{j=0}^{N_💡}(j+1)P_{j+1}^{(💡)}} = \frac{(k+1)P_{k+1}^{(💡)}}{\langle k\rangle_💡}.$$

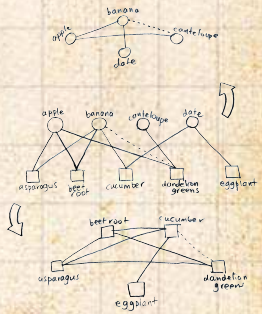


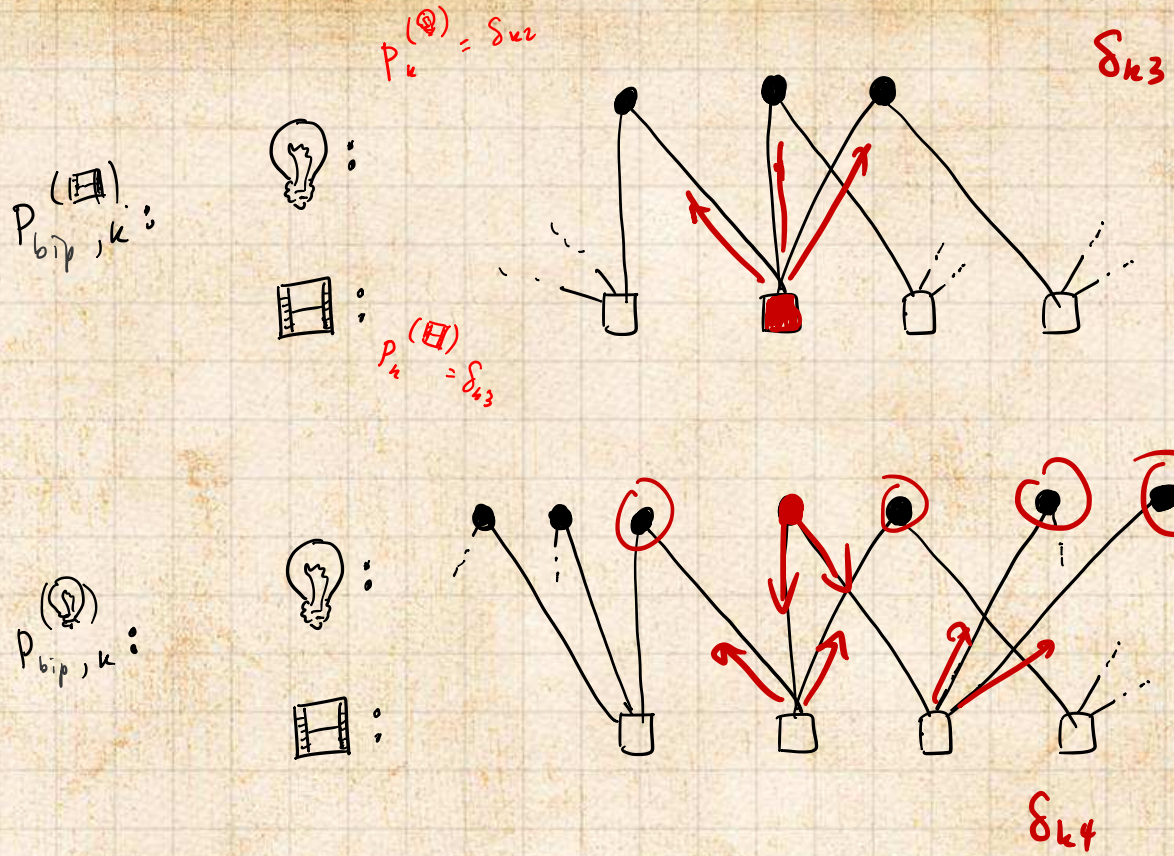
Networks of 🎞 and 💡 within bipartite structure:

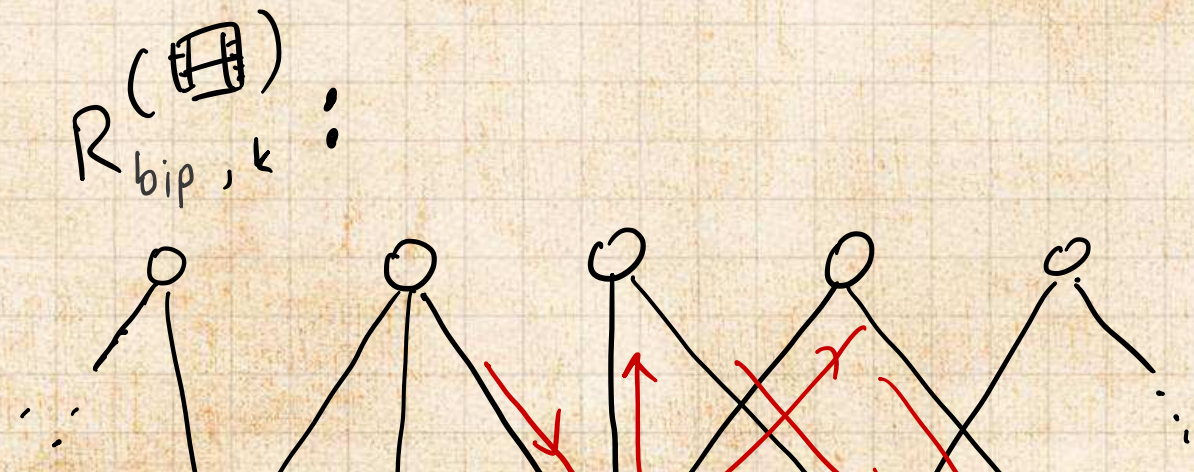
- $P^{(🎞)}_{\text{ind},k}$ = probability a random 🎞 is connected to k stories by sharing at least one 💡.
- $P^{(💡)}_{\text{ind},k}$ = probability a random 💡 is connected to k tropes by co-occurring in at least one 🎞.
- $R^{(💡—🎞)}_{\text{ind},k}$ = probability a random edge leads to a 🎞 which is connected to k other stories by sharing at least one 💡.
- $R^{(🎞—💡)}_{\text{ind},k}$ = probability a random edge leads to a 💡 which is connected to k other tropes by co-occurring in at least one 🎞.
- Goal: find these distributions 🔲.
- Another goal: find the induced distribution of component sizes and a test for the presence or absence of a giant component.
- Unrelated goal: be 10% happier/weep less.

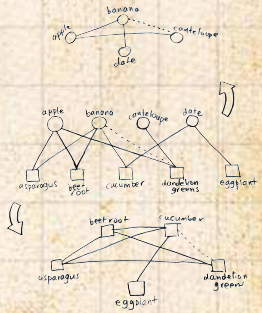


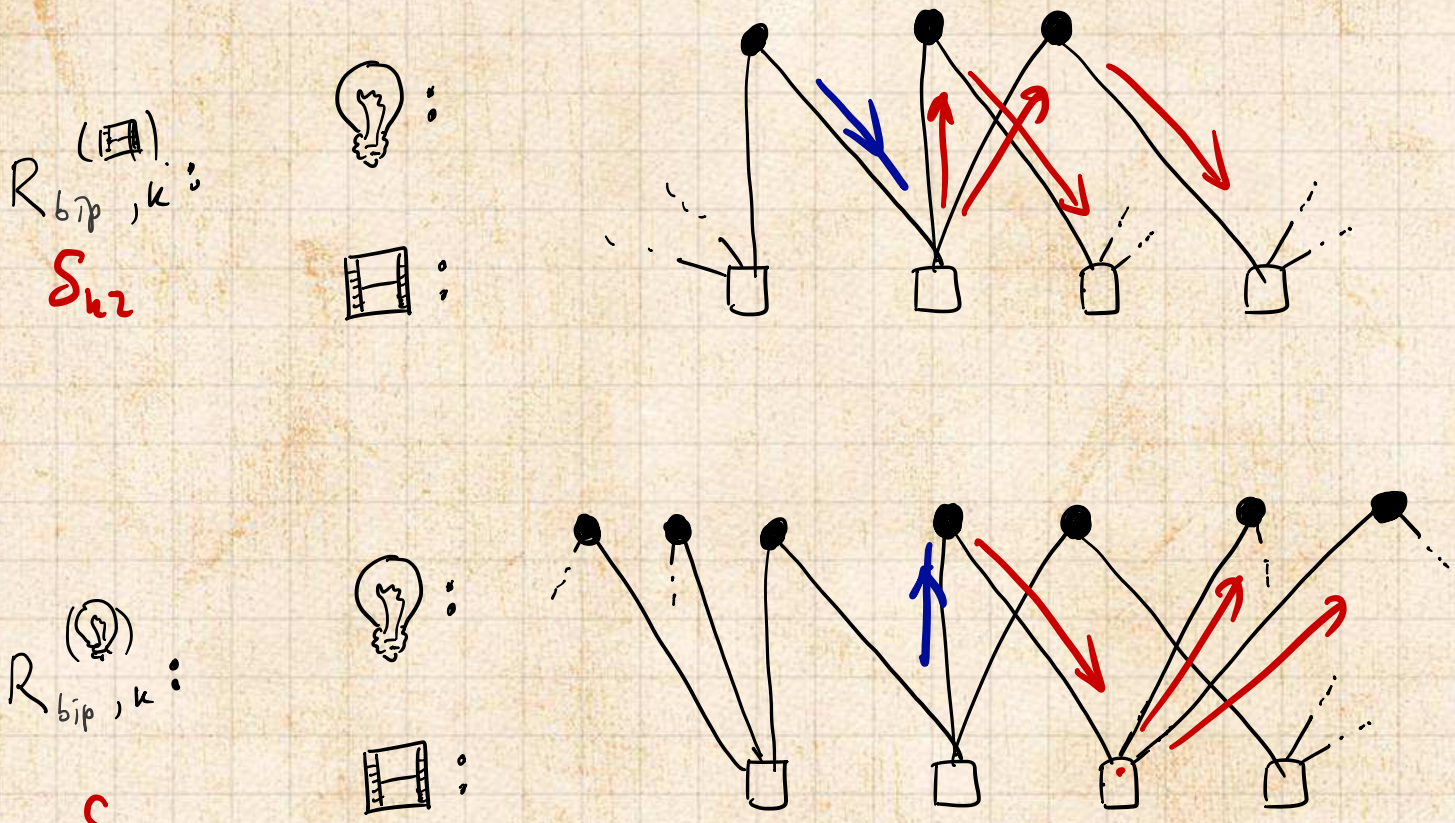

degree 4
in induced
network

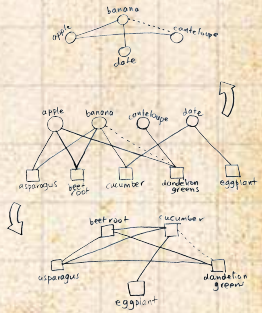

banana
apple
cantaloupe
date
apple
banana
cantaloupe
date
asparagus
beet root
cucumber
dandelion greens
eggplant
beet root
cucumber
asparagus
dandelion greens
eggplant


$P^{(\Omega)}_{k} = \delta_{k2}$
δ_{k3}
$P^{(\Omega)}_{bip,k}$:
$P^{(\Omega)}_{k} = \delta_{k3}$
$P^{(\Omega)}_{bip,k}$:
δ_{k4}


$R^{(\Omega)}_{bip,k}$:


banana
apple
cantaloupe
date
apple
banana
cantaloupe
date
asparagus
beet root
cucumber
dandelion greens
eggplant
beet root
cucumber
asparagus
dandelion greens
eggplant


$R_{bip,k}^{(\square)}$:
δ_{k2}
$R_{bip,k}^{(\Omega)}$:
δ_{k2}


banana
apple
cantaloupe
date
asparagus
beet root
cucumber
dandelion greens
eggplant

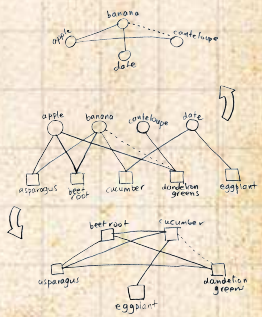
Generating Function Madness

Yes, we're doing it:

- $F_{P(\text{film})}(x) = \sum_{k=0}^{\infty} P_k^{(\text{film})} x^k$
- $F_{P(\text{bulb})}(x) = \sum_{k=0}^{\infty} P_k^{(\text{bulb})} x^k$
- $F_{R(\text{film})}(x) = \sum_{k=0}^{\infty} R_k^{(\text{film})} x^k = \frac{F'_{P(\text{film})}(x)}{F'_{P(\text{film})}(1)}$
- $F_{R(\text{bulb})}(x) = \sum_{k=0}^{\infty} R_k^{(\text{bulb})} x^k = \frac{F'_{P(\text{bulb})}(x)}{F'_{P(\text{bulb})}(1)}$

The usual goodness:

- Normalization: $F_{P(\text{film})}(1) = F_{P(\text{bulb})}(1) = 1.$
- Means: $F'_{P(\text{film})}(1) = \langle k \rangle_{\text{film}}$ and $F'_{P(\text{bulb})}(1) = \langle k \rangle_{\text{bulb}}.$



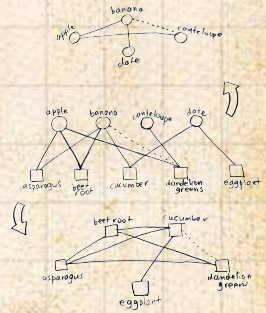
We strap these in as well:

- $F_{P^{(\text{film})}_{\text{ind}}}(x) = \sum_{k=0}^{\infty} P^{(\text{film})}_{\text{ind},k} x^k$
- $F_{P^{(\text{bulb})}_{\text{ind}}}(x) = \sum_{k=0}^{\infty} P^{(\text{bulb})}_{\text{ind},k} x^k$
- $F_{R^{(\text{bulb}-\text{film})}_{\text{ind}}}(x) = \sum_{k=0}^{\infty} R^{(\text{bulb}-\text{film})}_{\text{ind},k} x^k$
- $F_{R^{(\text{film}-\text{bulb})}_{\text{ind}}}(x) = \sum_{k=0}^{\infty} R^{(\text{film}-\text{bulb})}_{\text{ind},k} x^k$

So how do all these things connect?

- We're again performing sums of a randomly chosen number of randomly chosen numbers.
- We use one of our favorite sneaky tricks:

$$W = \sum_{i=1}^{U} V^{(i)} \rightleftharpoons F_W(x) = F_U(F_V(x)).$$



Induced distributions are not straightforward:

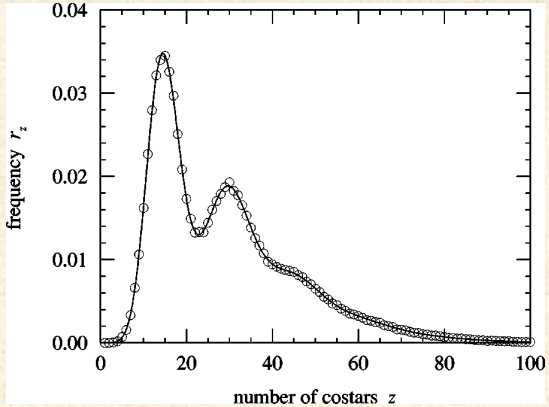
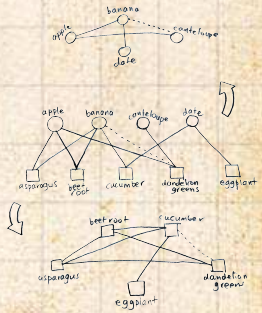
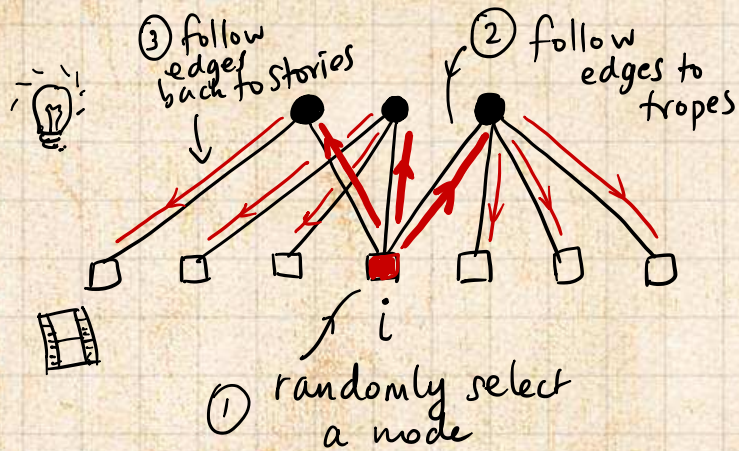


FIG. 7. The frequency distribution of numbers of co-stars of an actor in a bipartite graph with $\mu = 1.5$ and $\nu = 15$. The points are simulation results for $M = 10\,000$ and $N = 100\,000$. The line is the exact solution, Eqs. (89) and (90). The error bars on the numerical results are smaller than the points.

- View this as $P^{(🎞)}_{\text{ind},k}$ (the probability a story shares tropes with k other stories). [7]
- Result of purely random wiring with Poisson distributions for affiliation numbers.
- Parameters: $N_{🎞} = 10^4$, $N_{💡} = 10^5$, $\langle k \rangle_{🎞} = 1.5$, and $\langle k \rangle_{💡} = 15$.

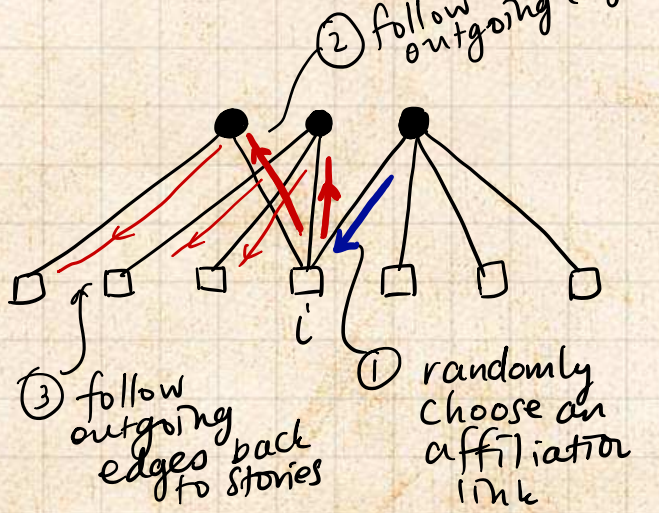


$$F_{p_{bip}}(\boxplus)(x) = F_p(\boxplus)\left(F_R^{(\lightbulb)}(x)\right)$$

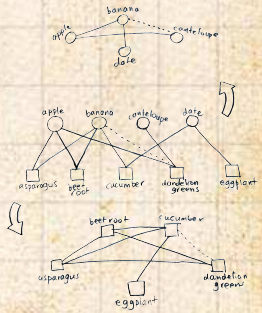


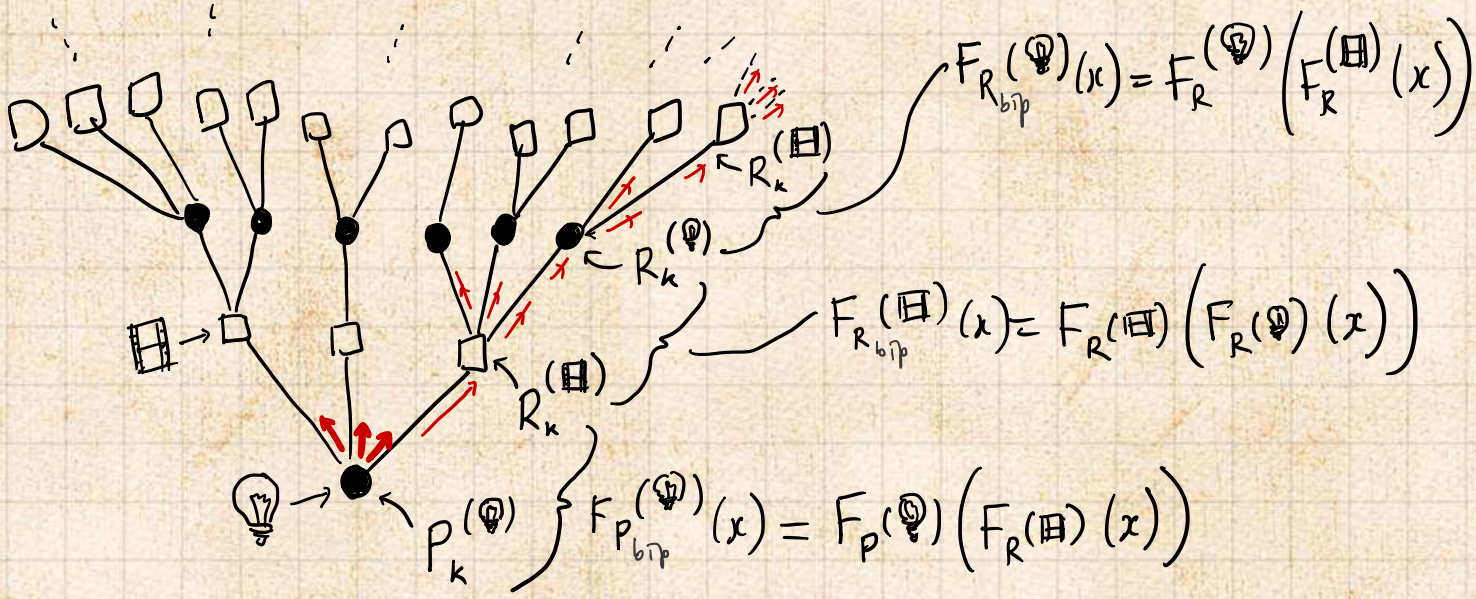
* i has 3 affiliations
* i has degree 6 in induced story network

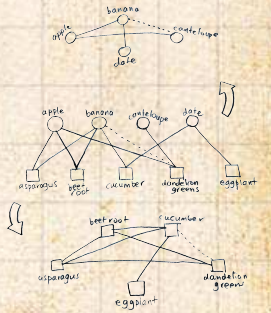
$$F_{R_{bip}}(\boxplus)(x) = F_R(\boxplus)\left(F_R^{(\lightbulb)}(x)\right)$$



* seems i has 3 outgoing edges
* but how depends on which edge we initially choose
* fine for distributions & gen. func. calculations




$F_{R_{bip}}^{(\lightbulb)}(x) = F_R^{(\lightbulb)}\left(F_R^{(\boxplus)}(x)\right)$
$R_k^{(\boxplus)}$
$R_k^{(\lightbulb)}$
$F_{R_{bip}}^{(\boxplus)}(x) = F_R^{(\boxplus)}\left(F_R^{(\lightbulb)}(x)\right)$
$R_k^{(\boxplus)}$
$P_k^{(\lightbulb)}$
$F_{P_{bip}}^{(\lightbulb)}(x) = F_P^{(\lightbulb)}\left(F_R^{(\boxplus)}(x)\right)$


banana
apple
cantaloupe
date
apple
banana
cantaloupe
date
asparagus
beet root
cucumber
dandelion greens
eggplant
beet root
cucumber
asparagus
dandelion greens
eggplant

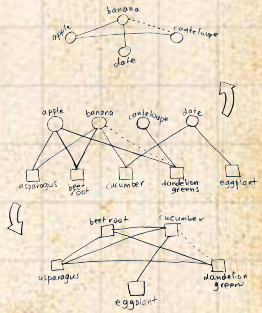
Induced distribution for stories:

- Randomly choose a 🎞, find its tropes (U), and then find how many other stories each of those tropes are part of (V):

$$F_{P^{(🎞)}_{\text{ind}}}(x) = F_{P^{(🎞)}_{\text{ind}}}(x) = F_{P^{(🎞)}}\left(F_{R^{(💡)}}(x)\right)$$

- Find the 🎞 at the end of a randomly chosen affiliation edge leaving a trope, find its number of other tropes (U), and then find how many other stories each of those tropes are part of (V):

$$F_{R^{(💡\text{–}🎞)}_{\text{ind}}}(x) = F_{R^{(🎞)}}\left(F_{R^{(💡)}}(x)\right)$$



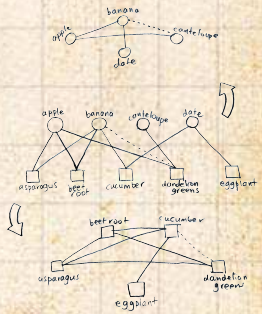
Induced distribution for tropes:

- Randomly choose a 💡, find the stories its part of (U), and then find how many other tropes are part of those stories (V):

$$F_{P^{(💡)}_{\text{ind}}}(x) = F_{P^{(💡)}_{\text{ind}}}(x) = F_{P^{(💡)}}\left(F_{R^{(🎞)}}(x)\right)$$

- Find the 💡 at the end of a randomly chosen affiliation edge leaving a story, find the number of other stories that use it (U), and then find how many other tropes are in those stories (V):

$$F_{R^{(🎞-💡)}_{\text{ind}}}(x) = F_{R^{(💡)}}\left(F_{R^{(🎞)}}(x)\right)$$



Let's do some good:

- Average number of stories connected to a story through trope-space:

$$\langle k \rangle_{🎞,\text{ind}} = F'_{P_{\text{ind}}^{(🎞)}}(1)$$

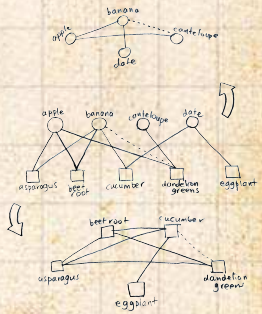
- So: $\langle k \rangle_{🎞,\text{ind}} = \left. \frac{\mathrm{d}}{\mathrm{d}x} F_{P^{(🎞)}}\left(F_{R^{(💡)}}(x)\right) \right|_{x=1}$

$$= F'_{R^{(💡)}}(1) F'_{P^{(🎞)}}\left(F_{R^{(💡)}}(1)\right) = F'_{R^{(💡)}}(1) F'_{P^{(🎞)}}(1)$$

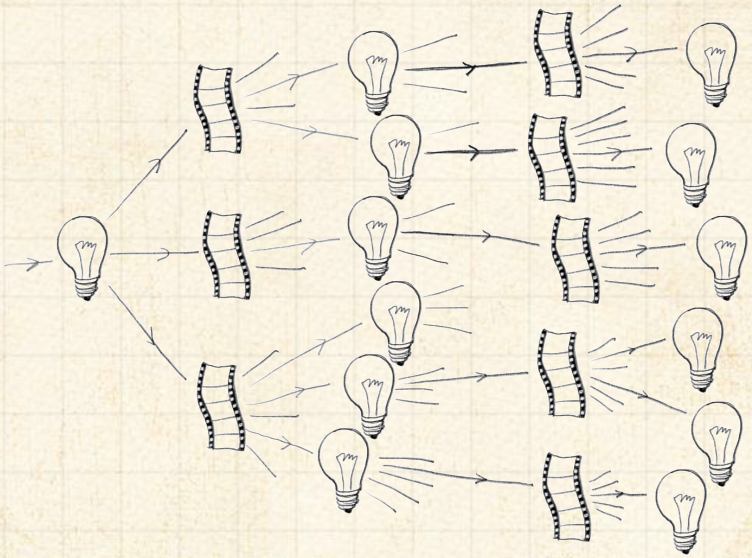
- Similarly, the average number of tropes connected to a random trope through stories:

$$\langle k \rangle_{💡,\text{ind}} = F'_{R^{(🎞)}}(1) F'_{P^{(💡)}}(1)$$

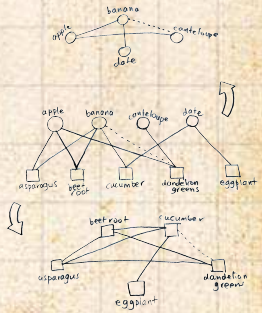
- In terms of the underlying distributions, we have: $\langle k \rangle_{🎞,\text{ind}} = \frac{\langle k(k-1) \rangle_{💡}}{\langle k \rangle_{💡}} \langle k \rangle_{🎞}$ and $\langle k \rangle_{💡,\text{ind}} = \frac{\langle k(k-1) \rangle_{🎞}}{\langle k \rangle_{🎞}} \langle k \rangle_{💡}$



Spreading through bipartite networks:



- View as bouncing back and forth between the two connected populations. [2]
- Actual spread may be within only one population (ideas between between people) or through both (failures in physical and communication networks).
- The gain ratio for simple contagion on a bipartite random network = product of two gain ratios.



Unstoppable spreading: is this thing connected?

- Always about the edges: when following a random edge toward a 🎞, what's the expected number of new edges leading to other stories via tropes?
- We want to determine $\langle k \rangle_{R,🎞,\text{ind}} = F'_{R^{(💡–🎞)}_{\text{ind}}}(1)$ (and $F'_{R^{(🎞–💡)}_{\text{ind}}}(1)$ for the trope side of things).
- We compute with joy:

$$\langle k \rangle_{R,🎞,\text{ind}} = \frac{\mathrm{d}}{\mathrm{d}x} F_{R^{(💡–🎞)}_{\text{ind},k}}(x)\bigg|_{x=1} = \frac{\mathrm{d}}{\mathrm{d}x} F_{R^{(🎞)}}\left(F_{R^{(💡)}}(x)\right)\bigg|_{x=1}$$

$$= F'_{R^{(💡)}}(1) F'_{R^{(🎞)}}\left(F_{R^{(💡)}}(1)\right) = F'_{R^{(💡)}}(1) F'_{R^{(🎞)}}(1) = \frac{F''_{P^{(💡)}}(1)}{F'_{P^{(💡)}}(1)} \frac{F''_{P^{(🎞)}}(1)}{F'_{P^{(🎞)}}(1)}$$

- Note symmetry.
- $happiness++;

- In terms of the underlying distributions:

$$\langle k \rangle_{R,🎞,\text{ind}} = \frac{\langle k(k-1) \rangle_{🎞}}{\langle k \rangle_{🎞}} \frac{\langle k(k-1) \rangle_{💡}}{\langle k \rangle_{💡}}$$

- We have a giant component in both induced networks when

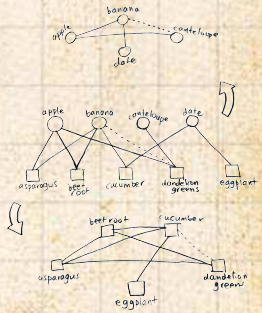
$$\langle k \rangle_{R,🎞,\text{ind}} \equiv \langle k \rangle_{R,💡,\text{ind}} > 1$$

.

- See this as the product of two gain ratios. #excellent #physics

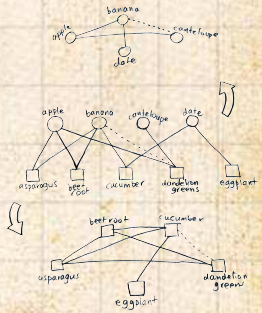
- We can mess with this condition to make it mathematically pleasant and pleasantly inscrutable:

$$\sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} kk'(kk' - k - k') P_k^{(🎞)} P_{k'}^{(💡)} = 0.$$



Simple example for finding the degree distributions for the two induced networks in a random bipartite affiliation structure:

- Set $P_k^{(🎞)} = \delta_{k3}$ and leave $P_k^{(💡)}$ arbitrary.
- Each story contains exactly three tropes.
- We have $F_{P^{(🎞)}}(x) = x^3$ and $F_{R^{(🎞)}}(x) = x^2$.
- Using $F_{P_{\text{ind}}^{(🎞)}}(x) = F_{P^{(🎞)}}\left(F_{R^{(💡)}}(x)\right)$ and $F_{P_{\text{ind}}^{(💡)}}(x) = F_{P^{(💡)}}\left(F_{R^{(🎞)}}(x)\right)$ we have $F_{P_{\text{ind}}^{(🎞)}}(x) = \left[F_{R^{(💡)}}(x)\right]^3$ and $F_{P_{\text{ind}}^{(💡)}}(x) = F_{P^{(💡)}}\left(x^2\right)$.
- Even more specific: If each trope is found in exactly two stories then $F_{P^{(💡)}} = x^2$ and $F_{R^{(💡)}} = x$ giving $F_{P_{\text{ind}}^{(🎞)}}(x) = x^3$ and $F_{P_{\text{ind}}^{(💡)}}(x) = x^4$.
- Yes for giant components ❑: $\langle k \rangle_{R,🎞,\text{ind}} \equiv \langle k \rangle_{R,💡,\text{ind}} = 2 \cdot 1 = 2 > 1$.



Boards and Directors: [7]

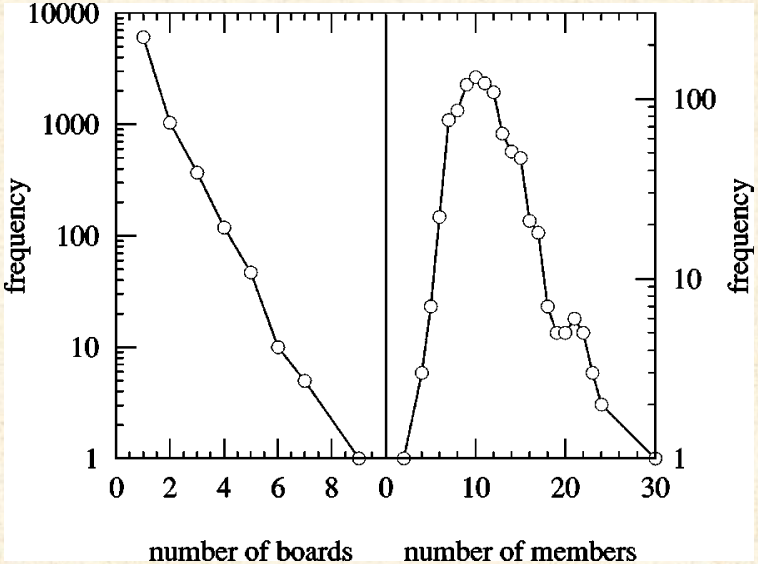
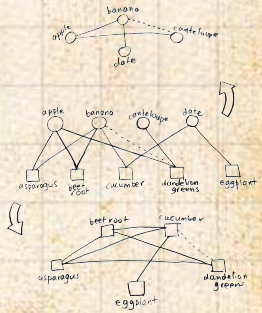


FIG. 8. Frequency distributions for the boards of directors of the Fortune 1000. Left panel: the numbers of boards on which each director sits. Right panel: the numbers of directors on each board.

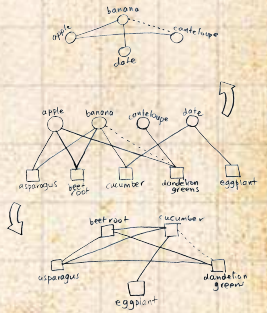
- Exponentialish distribution for number of boards each director sits on.
- Boards typically have 5 to 15 directors.
- Plan: Take these distributions, presume random bipartite structure and generate co-director network and board interlock network.



Boards and Directors and more: [7]

TABLE I. Summary of results of the analysis of four collaboration networks.

Network	Clustering C		Average degree z	
	Theory	Actual	Theory	Actual
Company directors	0.590	0.588	14.53	14.44
Movie actors	0.084	0.199	125.6	113.4
Physics (arxiv.org)	0.192	0.452	16.74	9.27
Biomedicine (MEDLINE)	0.042	0.088	18.02	16.93



- Random bipartite affiliation network assumption produces decent matches for some basic quantities.

Boards and Directors: [7]

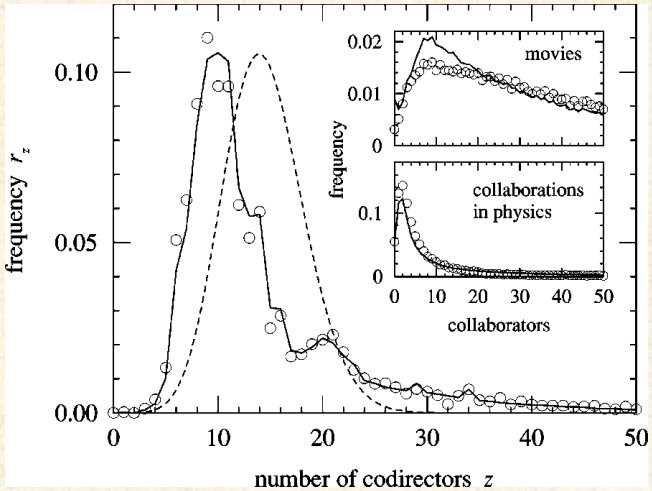
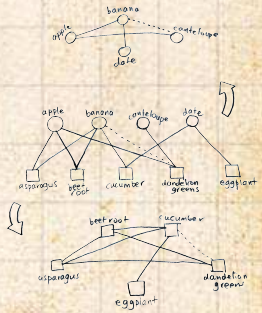


FIG. 9. The probability distribution of numbers of co-directors in the Fortune 1000 graph. The points are the real-world data, the solid line is the bipartite graph model, and the dashed line is the Poisson distribution with the same mean. Insets: the equivalent distributions for the numbers of collaborators of movie actors and physicists.

- Jolly good: Works very well for co-directors.
- For comparison, the dashed line is a Poisson with the empirical average degree.



Boards and Directors: [7]

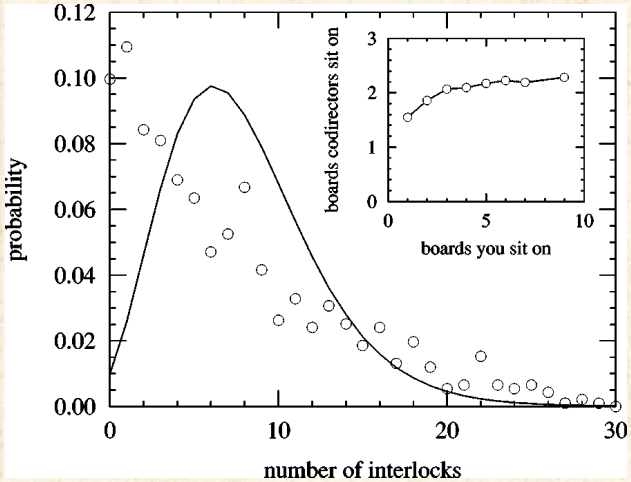
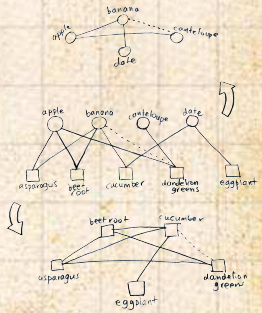


FIG. 10. The distribution of the number of other boards with which each board of directors is "interlocked" in the Fortune 1000 data. An interlock between two boards means that they share one or more common members. The points are the empirical data, the solid line is the theoretical prediction. Inset: the number of boards on which one's codirectors sit, as a function of the number of boards one sits on oneself.



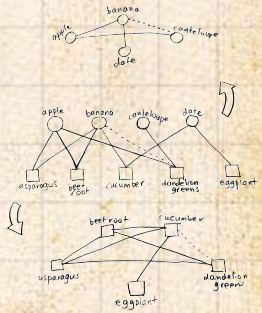
- Wins less bananas for the board interlock network.
- Assortativity is the reason: Directors who sit on many boards tend to sit on the same boards.
- Note: The term assortativity was not used in this 2001 paper.

To come:

- Distributions of component size.
- Simpler computation for the giant component condition.
- Contagion.
- Testing real bipartite structures for departure from randomness.

Nutshell:

- Random bipartite networks model many real systems well.
- Crucial improvement over simple random networks.
- We can find the induced distributions and determine connectivity/contagion condition.

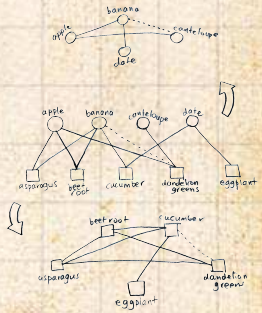


References I

[1] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási.
Flavor network and the principles of food pairing.
Nature Scientific Reports, 1:196, 2011. pdf

[2] P. S. Dodds.
A simple person's approach to understanding the contagion condition for spreading processes on generalized random networks.
In S. Lehmann and Y.-Y. Ahn, editors, Spreading Dynamics in Social Systems. 2017. pdf

[3] L. P. García-Pérez, M. A. Serrano, and M. Boguñá.
The complex architecture of primes and natural numbers, 2014.
http://arxiv.org/abs/1402.3612. pdf

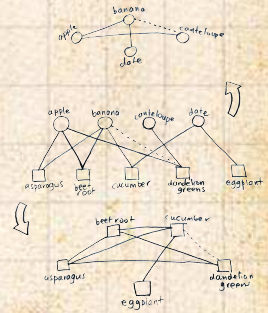


References II

[4] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási.
The human disease network.
Proc. Natl. Acad. Sci., 104:8685–8690, 2007. pdf

[5] R. Guimerà, B. Uzzi, J. Spiro, and L. A. N. Amaral.
Team assembly mechanisms determine collaboration network structure and team performance.
Science, 308:697–702, 2005. pdf

[6] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausman.
The product space conditions the development of nations.
Science, 317:482–487, 2007. pdf



References III

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts.
Random graphs with arbitrary degree distributions and their applications.
Phys. Rev. E, 64:026118, 2001. pdf

[8] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic.
Recipe recommendation using ingredient networks.
In Proceedings of the 3rd Annual ACM Web Science Conference, WebSci '12, pages 298–307, New York, NY, USA, 2012. ACM. pdf