netherlands eScience center

A First Course in Network Science

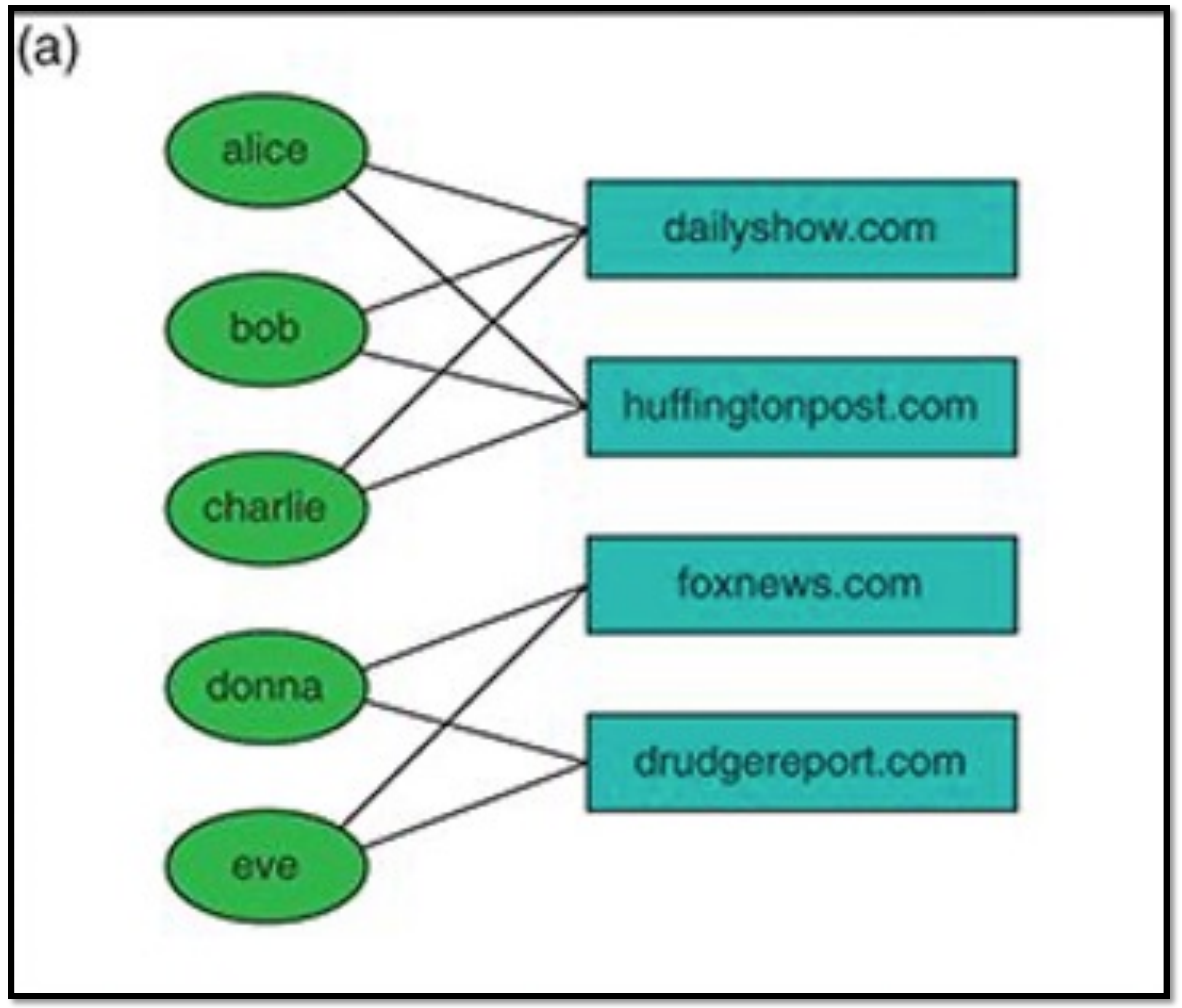
Chapter 4.6-4.7

Cunliang Geng
Amsterdam, 12/04/2021

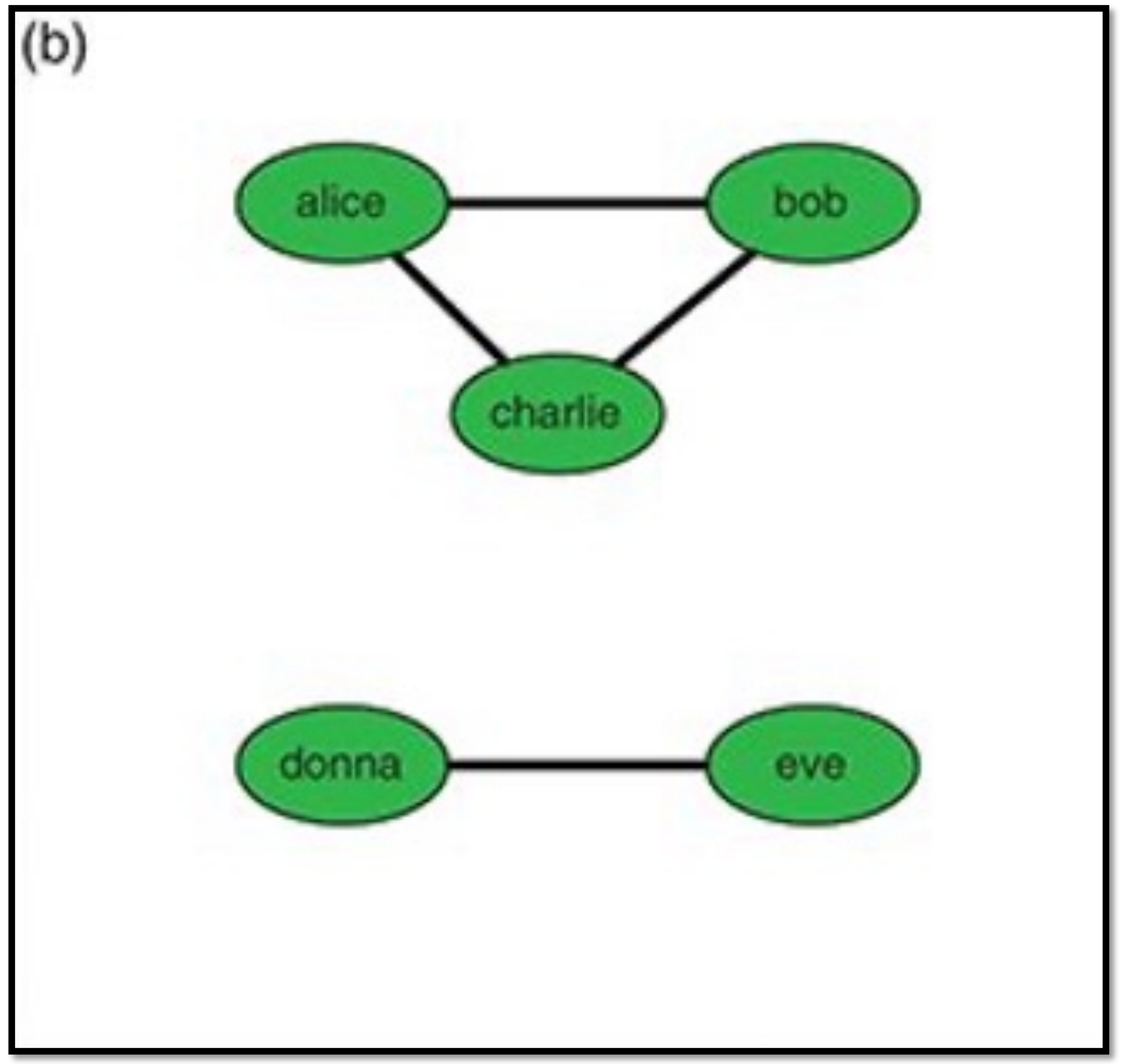
Recall: chapter 4

Co-occurrence Network

Unweighted bipartite network



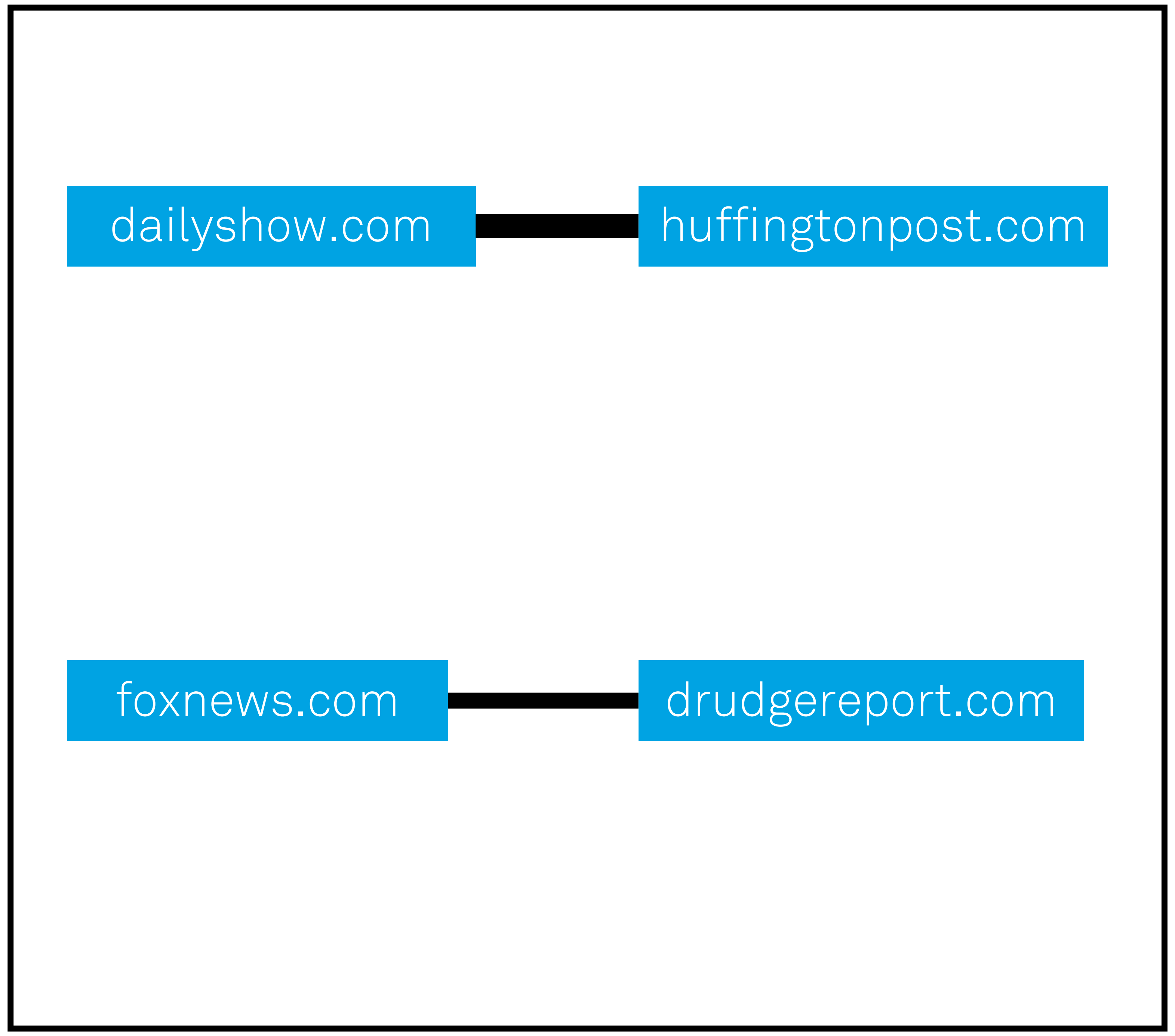
(a) A unweighted bipartite network induced by “like” relationships.



(b) A user co-occurrence network derived from projecting the “like” network onto user nodes.

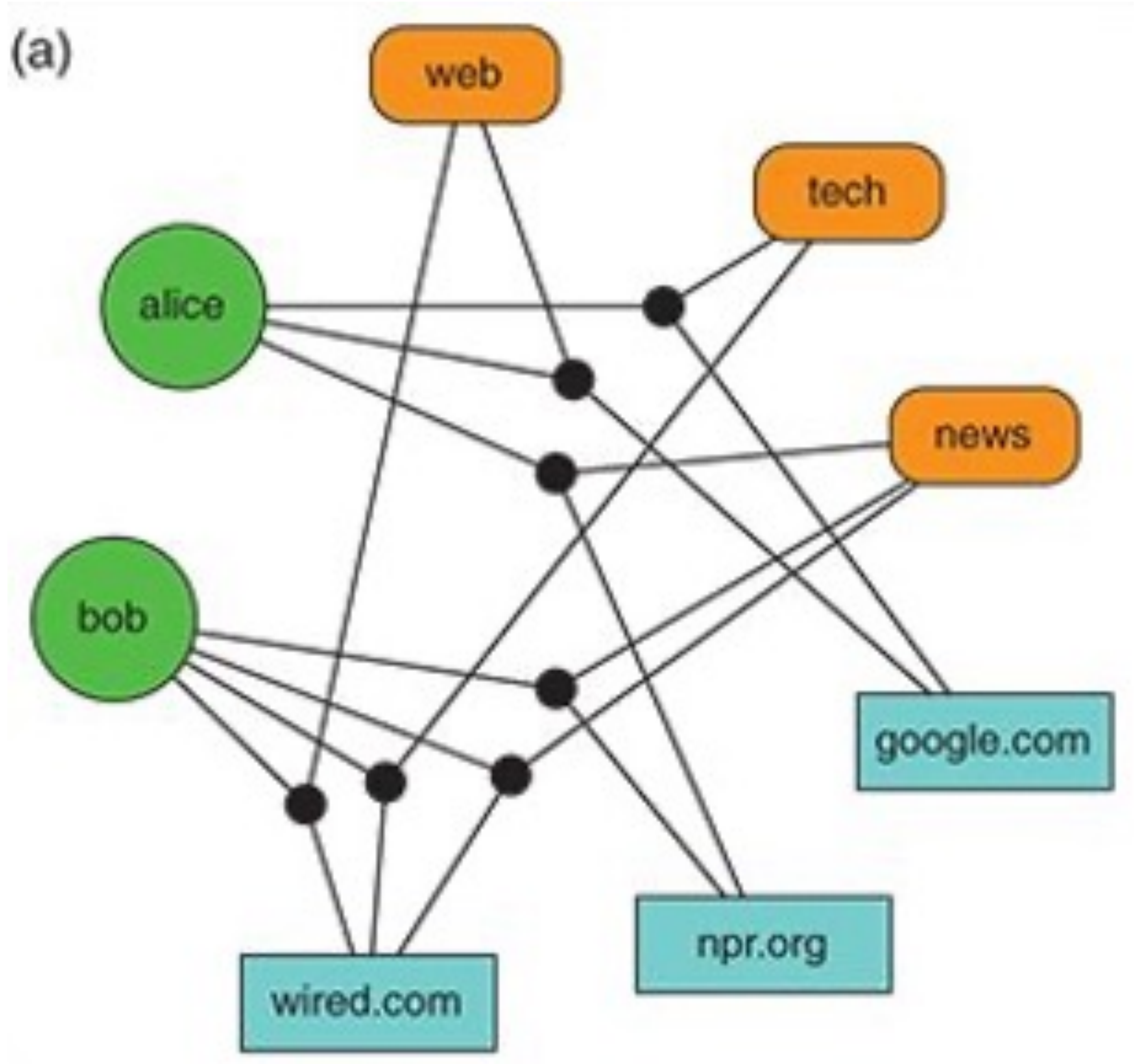
Question:

What does the website co-occurrence network look like?

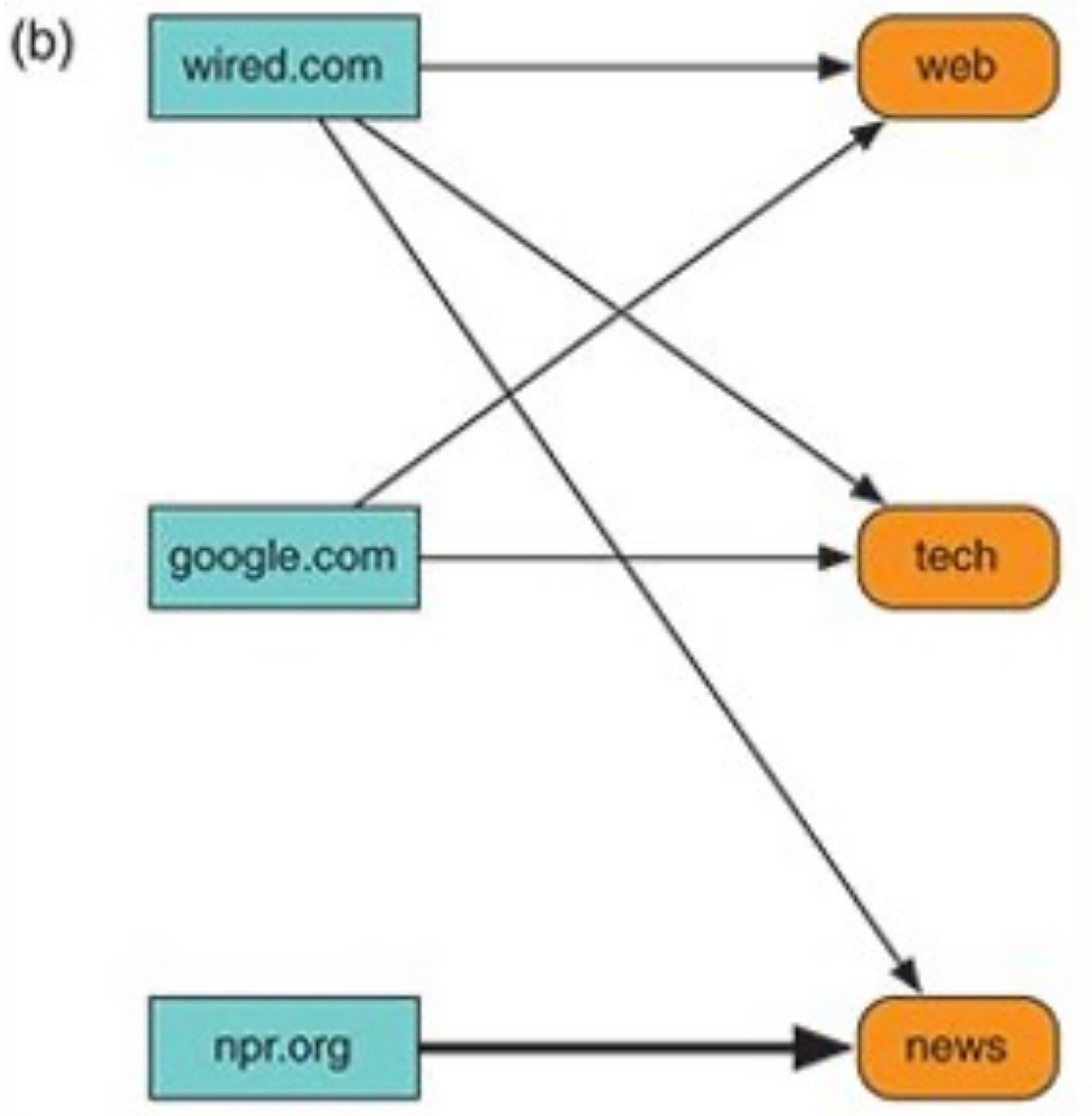


Co-occurrence Network

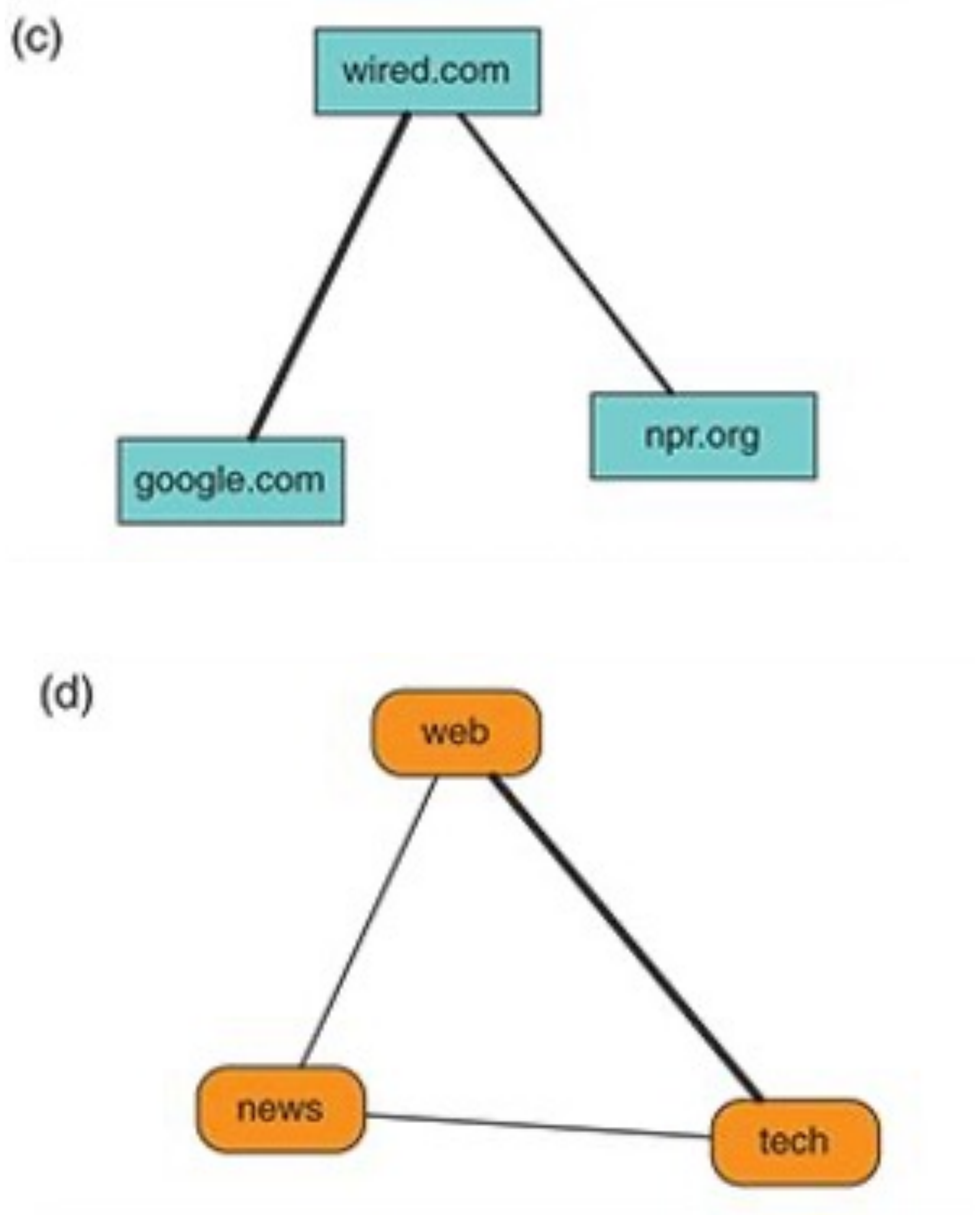
Weighted bipartite network: social tagging



Folksonomy
= folk + taxonomy



Weighted bipartite network



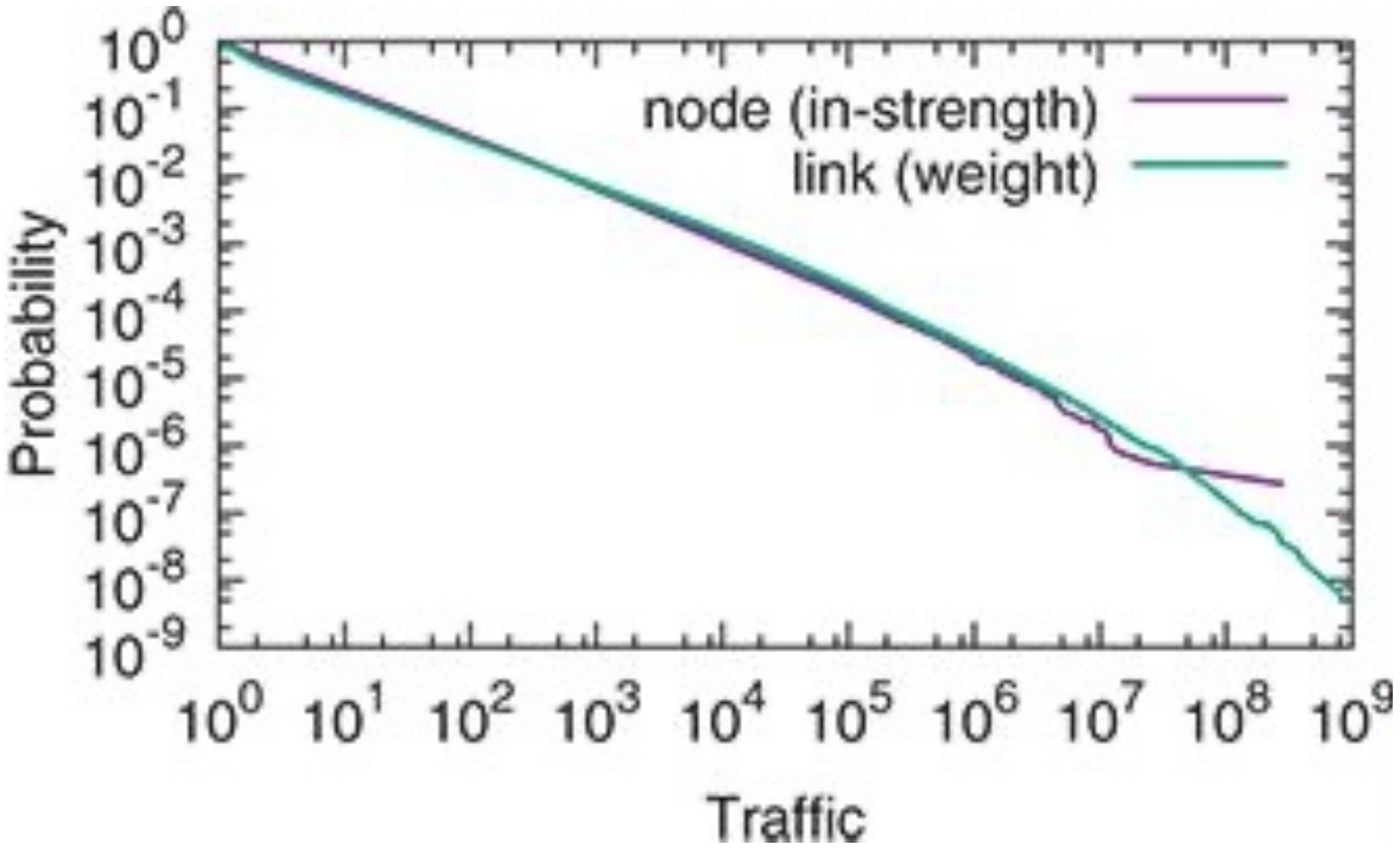
Co-occurrence networks for websites (C) and tags (D)

Questions:

How to preserve the information about the weights of bipartite network in the co-occurrence networks?

Weight Heterogeneity

Heavy tail

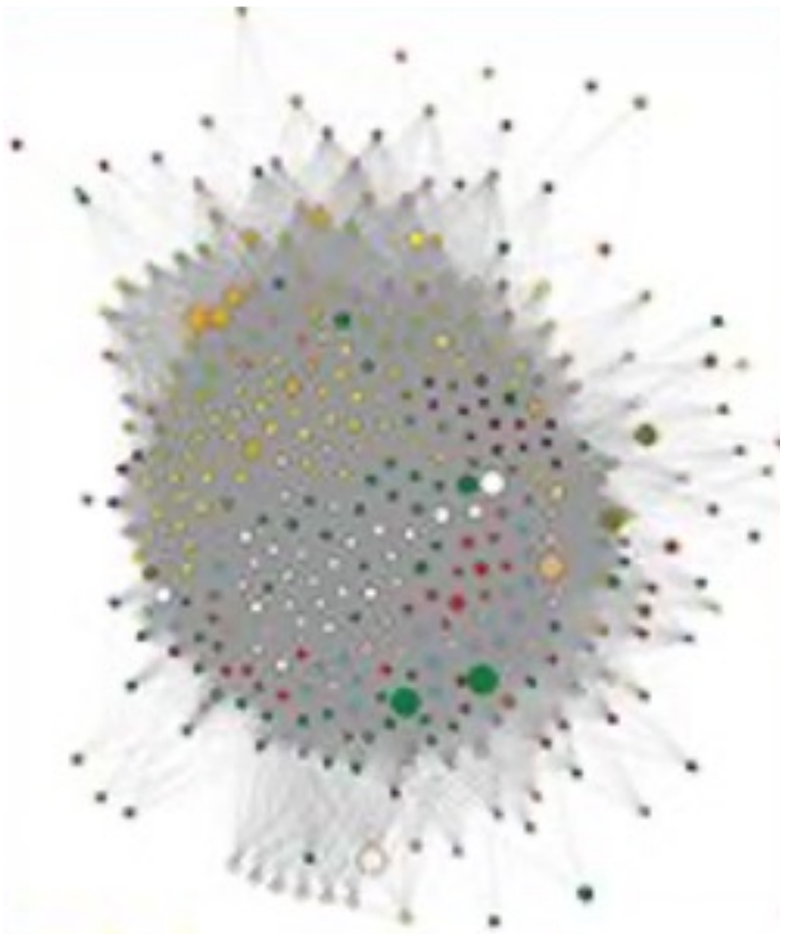


The distribution of website traffic (total number of clicks to a site), expressed by node in-strength, and of link traffic (total number of clicks on a hyperlink), expressed by link weight.

Weight Heterogeneity

Link filtering

Dense networks are hard to visualize.
So it is helpful to prune low-weight links in a weighted network.



How to prune a dense network?

A global weight threshold

Bad method.
Due to the heavy-tailed weight distribution, the majority of the nodes will be isolated.

Different thresholds for different nodes

e.g. keep 10% of links with the largest weights for each node

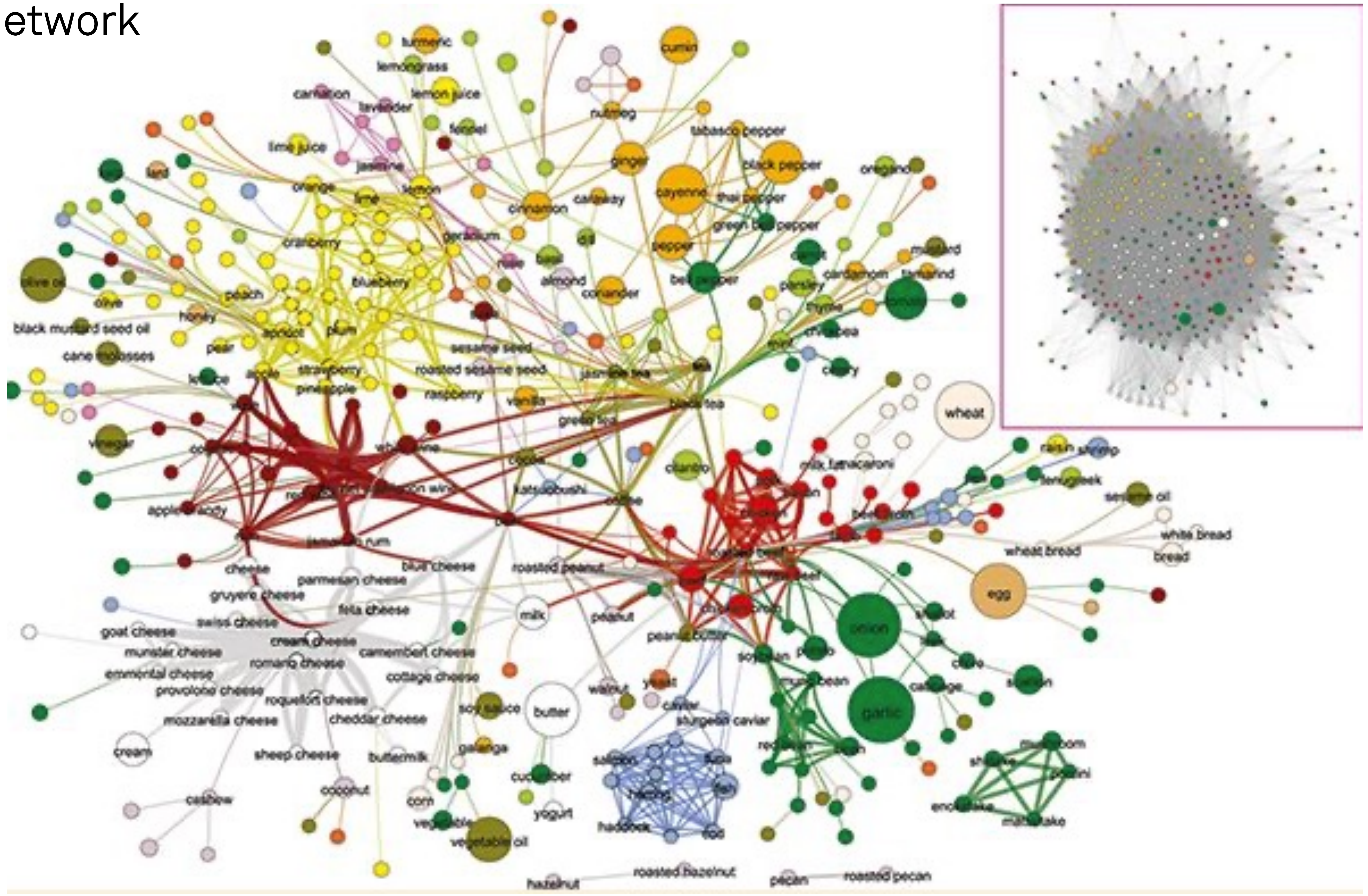
Not great. Not sure if it will keep all significant links or keep some non-sigificant links.

To identify the statistically significant links for each node,

a more principled approach is to find the network backbone,
i.e. detect the links that carry a disproportionate fraction of each node's strength.

Weight Heterogeneity

An example of backbone network



A flavor network:
Each node denotes an ingredient, its color indicates a food category, and its size reflects the ingredient prevalence in recipes.
Two ingredients are connected if they share flavor compounds, with the link width representing the number of shared compounds.
The full network is shown in the inset.

Summary

1. Co-occurrence networks often result from bipartite graphs.

2. The meaning of co-occurrence network: the weight of a link between two nodes a and b of the same type is a measure of how many nodes of another type are associated with both a and b.

3. Weighted networks can be very dense and heterogeneous, such as often having heavy-tailed weight distributions.

4. Prune a dense network is to extract the backbone of a heterogeneous weighted network by identifying the statistically significant links of each node.

Thank you!
Q&A

Weight Heterogeneity

How to find the network backbone?

In networks with broad distributions of link weights, using a global threshold to prune links is inappropriate. Instead, we can use the weight fluctuations for each node to identify the links to be preserved — those that carry most of the weight. Given a node i with degree k_i and strength s_i, let us evaluate a link against a null model in which the weights are distributed randomly on the k_i links adjacent to i, with the constraint that their sum equals s_i. The probability that a link has weight w_{ij} or larger under this hypothesis is

$$p_{ij} = \left(1 - \frac{w_{ij}}{s_i}\right)^{k_i - 1}. \qquad (4.2)$$

So, if link ij has weight w_{ij}, from Eq. (4.2) we compute the probability p_{ij} that such a value is compatible with the null model: if $p_{ij} < \alpha$, where α is a parameter that represents the desired significance level, the link is preserved, otherwise it is removed. Lower values of α lead to sparser networks, as fewer links are preserved. Since a link is connected to two nodes, we can obtain two values for p_{ij} by plugging the strength and degree of either node into Eq. (4.2). We can then use the larger or smaller of these values, depending on whether we wish to prune more or less aggressively. This link ltering procedure extracts a *network backbone*, which is supposed to preserve the essential structure and global properties of the network.