Are graphs the right model?

- Coauthor graph
  - $vw \in E(G)$ if author v and author w have written a paper together.

    — single author papers?

    — is a triangle the result of one 3-author paper or three 2-author papers...
- Movie Actor Graph.
  - lots of edges. One movie with 100 actors $\Rightarrow \binom{100}{2} = 4950$ edges

    $vw \in E(G)$ if actor v and actor w have appeared in a movie together. — we lost most of the information about movies. to recover it we have to look for large maximal cliques.
- Things like Facebook groups
  - If we use "groups graph" then huge groups generate too many edges

    Take friend graph and find communities, you may discover that communities highly correlate with group membership.
- Some facebook groups have 5M members. $\Rightarrow 1.2 \times 10^{12}$ edges!

  Community finding: One drawback of community-finding algorithms is that they insist on communities being disjoint. Usually this isn't true.

Hypergraphs

Represent multi-way relationships
- coauthors of a paper
- all actors in a movie

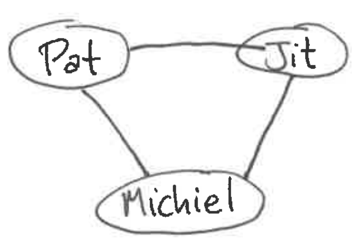
  ⋮

$H = (V(H), E(H))$

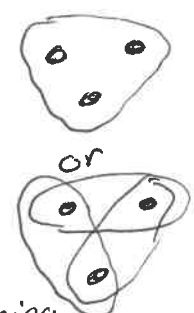
$V(H)$ is a set of vertices / nodes

$E(H)$ is a set of hyperedges. Each element in $E(H)$ is a subset of $V(H)$.

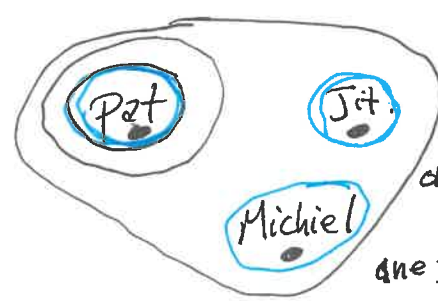
Advantages: Hypergraphs retain information that gets lost when we use a graph model.



Did these three authors write one 3-author paper or three 2-author papers?



- Even edges of size 1 have meaning



Pat has one single-author paper and one 3-author paper.

Disadvantages:
- Hard to draw!
- The theory of hypergraphs is less developed than graph theory
- Usually more than one version of any graph theory problem.

- $\deg_H(v) = |H(v)|$ is the number of hyperedges that contain v.

- ~~Ins~~ $H(v)$ is like $N_G(v)$

$|H(v)| = \deg_H(v)$ is like $\deg_G(v) = |N_G(v)|$

$\Delta(H) = \max\{\deg_H(v) : v \in V(H)\}$ = max-degree(H)

$\delta(H) = \min\{\deg_H(v) : v \in V(H)\}$

H is <u>d-regular</u> if $\Delta(H) = \delta(H) = \deg_H(v)$ for each $v \in V(H)$.

$r(H) = \text{rank}(H) = \max\{|e| : e \in E(H)\}$

$cr(H) = \text{co-rank}(H) = \min\{|e| : e \in E(H)\}$

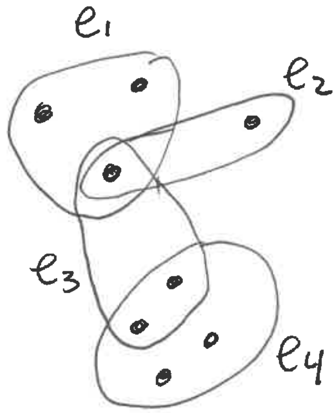
A <u>walk</u> in H is a sequence of vertices $w_0, w_1, \dots, w_r$ of H such that, for each $i \in \{1, \dots, r\}$, there exists an edge $e_i \in E(H)$ with $\{w_{i-1}, w_i\} \subseteq e_i$. Sometimes we can be more specific and write $w_0, e_1, w_1, e_2, w_2, \dots, w_{r-1}, e_r, w_r$. The second representation contains more information since H may have more than one edge that contains w_{i-1} and w_i

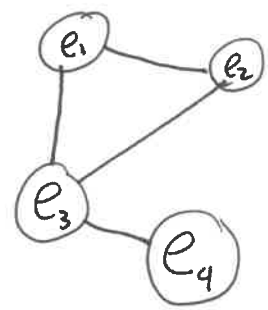
Line Graph $L(H)$ is a graph with

$$V(L(H)) = E(H).$$

$$E(L(H)) = \left\{ ef \in \binom{E(H)}{2} : e \cap f \neq \emptyset \right\}.$$

(ef is an edge of $L(H)$ iff e and f are incident)

H: 



: $L(H)$

Warning: We lose lots of information about $V(H)$.

Random Hypergraph Models

- Randomly-generated graph models are useful for testing and also for calibrating algorithms.

Binomial Random Hypergraph

- Let $\mathbf{p} = (p_1, \dots, p_K) \in [0,1]^K$ be a sequence of probabilities.
- $\mathcal{H}(n, \mathbf{p})$ has vertex set $V(\mathcal{H}(n,\mathbf{p})) = \{1, \dots, n\} = [n]$

  For each $i \in \{1, \dots, K\}$ and each $e \in \binom{[n]}{i}$,

  $e \in E(\mathcal{H}(n,\mathbf{p}))$ with probability p_i.

- Usually want p_i to decrease with n. Typically

$$p_i = c_i \Big/ \binom{n-1}{i-1}$$

then for any vertex $v \in [n]$

probability that each i-element subset becomes a hyperedge.

$$\mathbb{E}\left(\deg_{\mathcal{H}(n,\mathbf{p})}(v)\right) = \sum_{i=1}^{K} \underbrace{\binom{n-1}{i-1}}_{\substack{\#\ i\text{-element subsets}\\ \text{of } \{1,\dots,n\} \text{ that contain } v}} p_i = \sum_{i=1}^{K} c_i$$

Random Binomial K-uniform Hypergraph. $\mathcal{H}(n,p,k)$

- Use $\mathbf{p} = (\underbrace{0,0,\ldots,0}_{k-1}, p)$ $\quad p_i = 0$ for $i \in \{1 \ldots k-1\}$ $p_k = p$.

Typically $p = c / \binom{n-1}{k-1}$

Note: $\mathcal{H}(n,\mathbf{p}) = \bigcup_{i=1}^{k} \mathcal{H}(n,p_i,i)$ so studying $\mathcal{H}(n,p,k)$ isn't such a stupid idea. In fact, some of the properties of $\mathcal{H}(n,\mathbf{p})$ are mostly determined by $\mathcal{H}(n,p_i,i)$ for some critical value of i.

Basic Properties of $\mathcal{H}(n,p,k)$, when $p = c/\binom{n-1}{k-1}$

- If $c < 1/(k-1)$ then the largest component has size $O(\log n)$.
- If $c > 1/(k-1)$ then there is exactly one component of size $\Theta(n)$ and the other components have size $O(\log n)$.
- Hypergraph is connected when the minimum degree becomes 1, which happens when $c \geq \ln n + \omega(1)$.

All true with prob that goes to 1 as $n \to \infty$.