

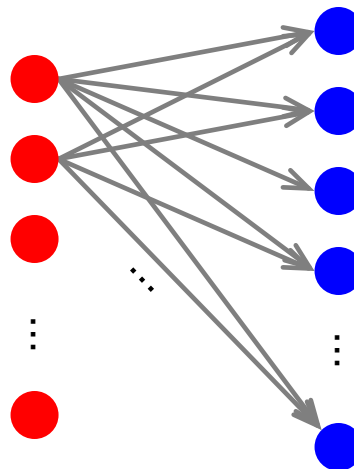
Analysis of Large Graphs: Trawling

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University



Trawling

- Searching for small communities in the Web graph
- What is the signature of a community / discussion in a Web graph?



Dense 2-layer graph

Use this to define “topics”:
What the same people on
the left talk about on the right
Remember HITS!

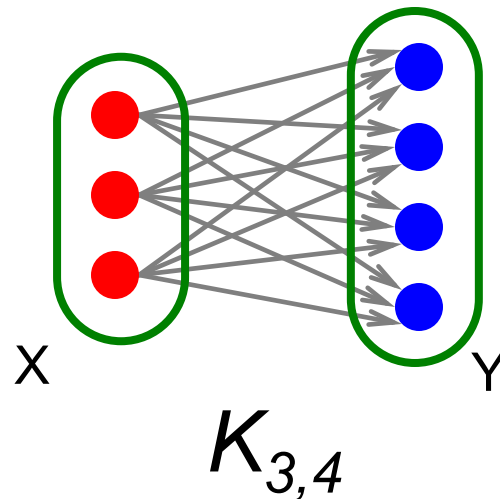
Intuition: Many people all talking about the same things

Searching for Small Communities

- **A more well-defined problem:**

Enumerate complete bipartite subgraphs $K_{s,t}$

- Where $K_{s,t}$: s nodes on the “left” where each links to the same t other nodes on the “right”



$$\begin{aligned} |X| &= s = 3 \\ |Y| &= t = 4 \end{aligned}$$

Fully connected

Frequent Itemset Enumeration

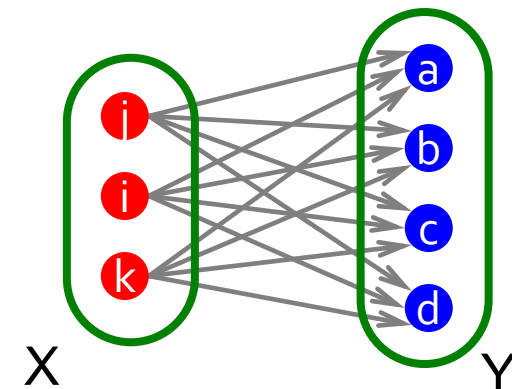
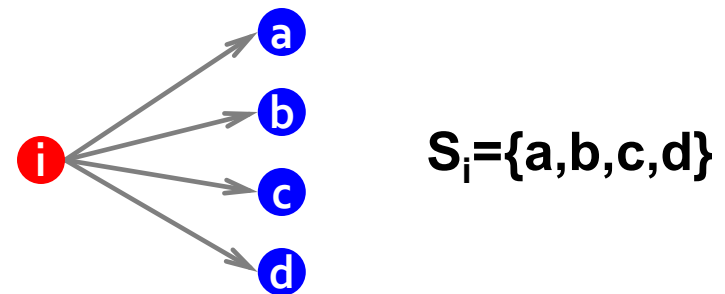
- **Market basket analysis.** Setting:
 - **Market:** Universe U of n items
 - **Baskets:** m subsets of U : $S_1, S_2, \dots, S_m \subseteq U$
(S_i is a set of items one person bought)
 - **Support:** Frequency threshold f
- **Goal:**
 - Find all subsets T s.t. $T \subseteq S_i$ of at least f sets S_i
(items in T were bought together at least f times)
- **What's the connection between the itemsets and complete bipartite graphs?**

From Itemsets to Bipartite $K_{s,t}$

Frequent itemsets = complete bipartite graphs!

■ How?

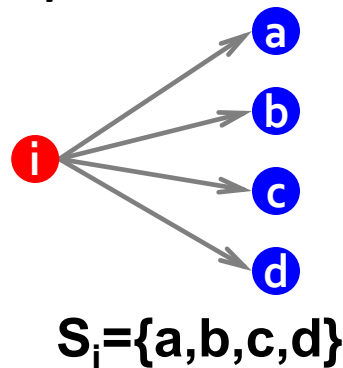
- View each node i as a set S_i of nodes i points to
- $K_{s,t}$ = a set Y of size t that occurs in s sets S_i
- Looking for $K_{s,t} \rightarrow$ set of frequency threshold to s and look at layer t – all frequent sets of size t



s minimum support ($|X|=s$)
 t itemset size ($|Y|=t$)

From Itemsets to Bipartite $K_{s,t}$

View each node i as a set S_i of nodes i points to

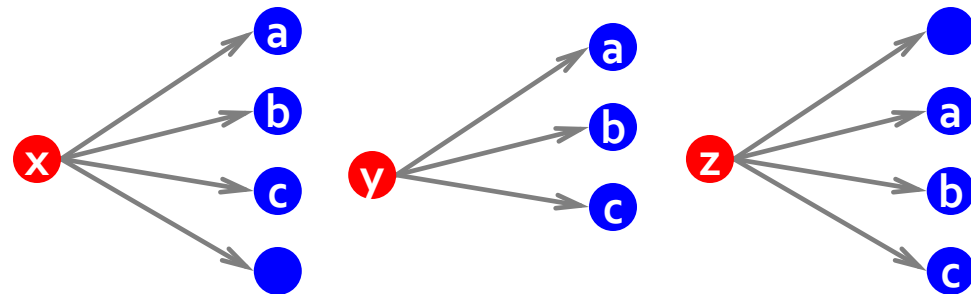


Find frequent itemsets:

s minimum support

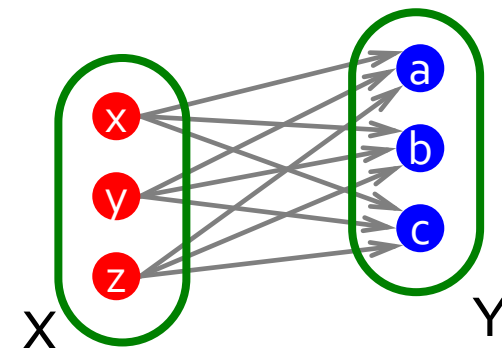
t itemset size

Say we find a **frequent itemset** $Y = \{a, b, c\}$ of supp s
 So, there are s nodes that link to all of $\{a, b, c\}$:

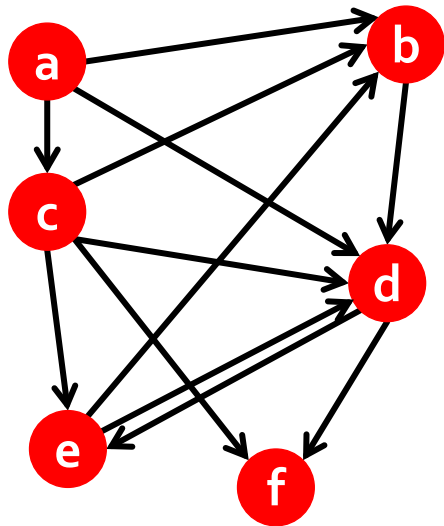


We found $K_{s,t}$!

$K_{s,t}$ = a set Y of size t
 that occurs in s sets S_i



Example (1)



Itemsets:

$a = \{b, c, d\}$

$b = \{d\}$

$c = \{b, d, e, f\}$

$d = \{e, f\}$

$e = \{b, d\}$

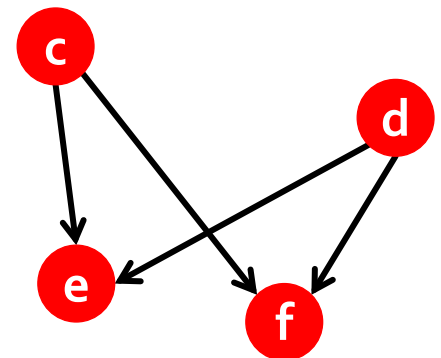
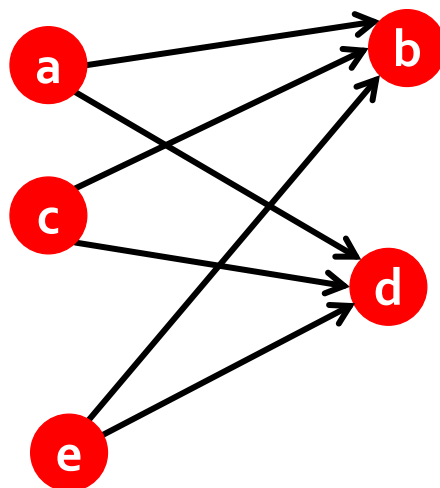
$f = \{\}$

- **Support threshold $s=2$**

- $\{b, d\}$: support 3

- $\{e, f\}$: support 2

- **And we just found 2 bipartite subgraphs:**



Example (2)

■ Example of a community from a web graph

A community of Australian fire brigades

Nodes on the right	Nodes on the left
NSW Rural Fire Service Internet Site NSW Fire Brigades Sutherland Rural Fire Service CFA: County Fire Authority “The National Cente...ted Children’s Ho... CRAFTI Internet Connexions-INFO Welcome to Blackwoo... Fire Safety Serv... The World Famous Guestbook Server Wilberforce County Fire Brigade NEW SOUTH WALES FIR...ES 377 STATION Woronora Bushfire Brigade Mongarlowe Bush Fire – Home Page Golden Square Fire Brigade FIREBREAK Home Page Guises Creek Volunt...fficial Home Page...	New South Wales Fir...ial Australian Links Feuerwehrlinks Australien FireNet Information Network The Cherrybrook Rur...re Brigade Home Page New South Wales Fir...ial Australian Links Fire Departments, F... Information Network The Australian Firefighter Page Kristiansand brannv...dens brannvesener... Australian Fire Services Links The 911 F,P,M., Fir...mp; Canada A Section Feuerwehrlinks Australien Sanctuary Point Rural Fire Brigade Fire Trails “l...ghters around the... FireSafe – Fire and Safety Directory Kristiansand Firede...departments of th...