



**ON COMMUNITIES: ALGORITHMS,
EXPERIMENTS, AND EXPERIENCES**

**Yi Chen
Computer Science Department
Pomona College**



○ facts:

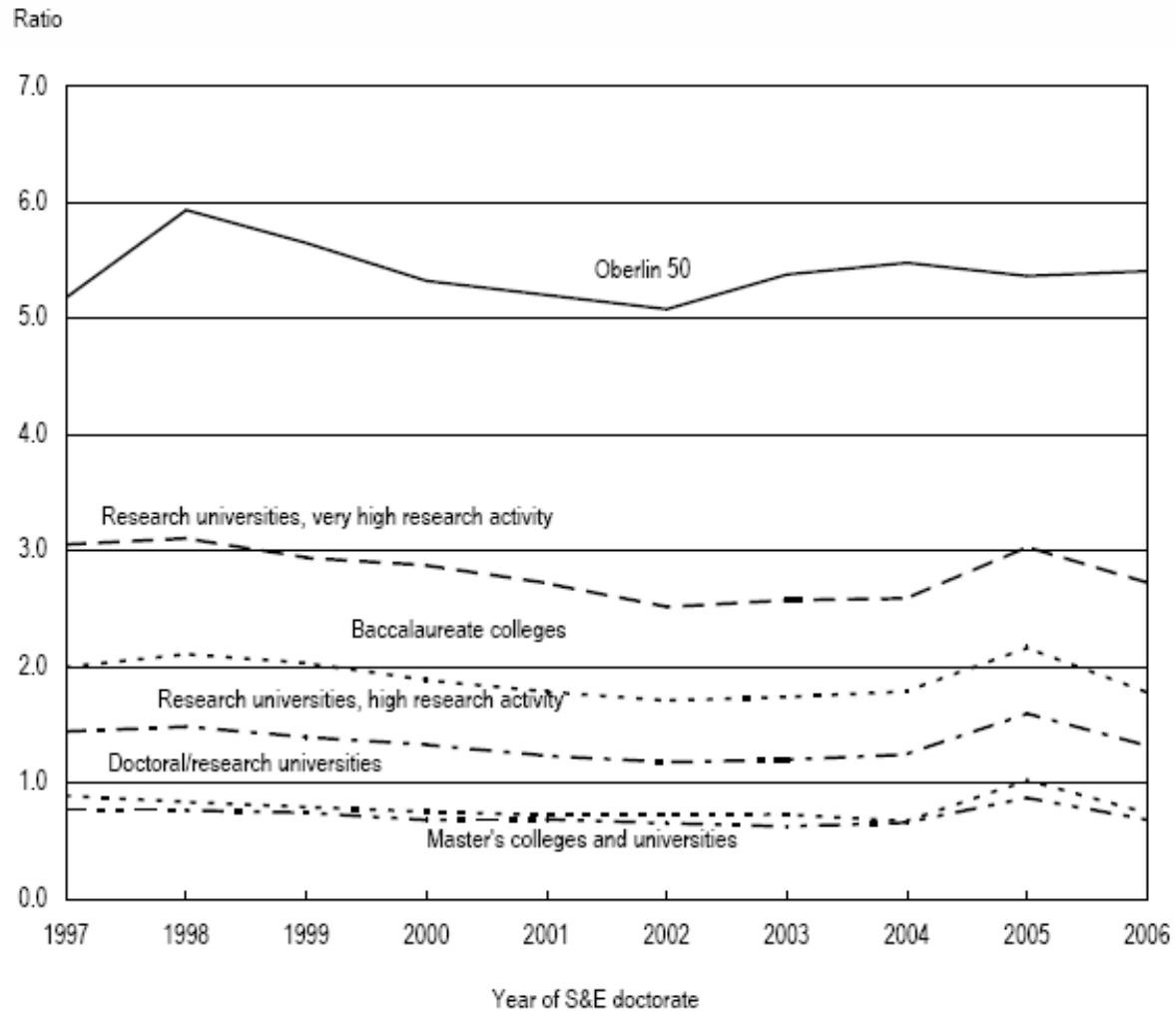
- ~1600 undergraduates, no graduate students
- part of Claremont Consortium (5 undergraduate + 2 graduate institutions located in adjacent campuses) located 35 miles East of Los Angeles.
- ~30 computer science majors graduate each year
- CS program had 1.5 faculty in 2002, CS department now has 6 tenure-track faculty (hiring again this year)

○ values:

- “They only are loyal to this college who, departing, bear their added riches in trust for mankind”
- loan free, strong commitment to diversity, emphasis on faculty-student relationship

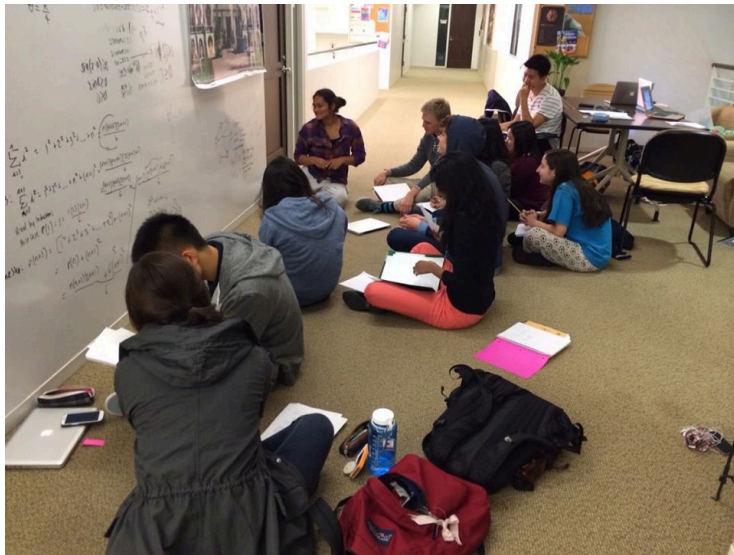


FIGURE 1. S&E doctorates per hundred bachelor's degrees awarded in all fields 9 years earlier, by 2005 Carnegie classification and Oberlin 50: 1997–2006



SOURCES: National Science Foundation, Division of Science Resources Statistics, Survey of Earned Doctorates, 1997–2006 and special tabulations of U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System, Completions Survey, 1988–1997.

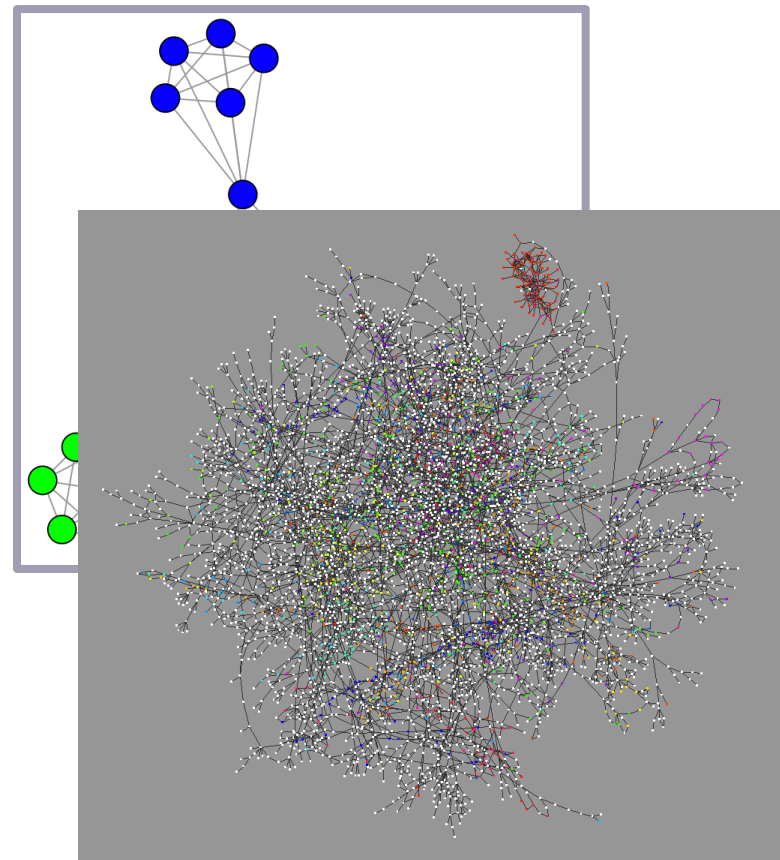
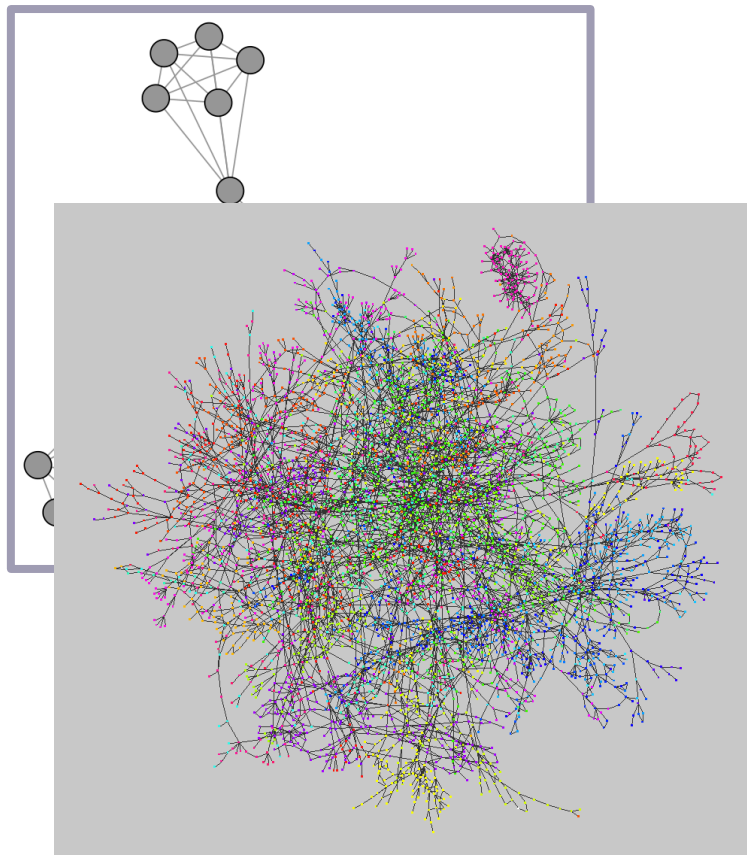
COMMUNITIES



MORE COMMUNITIES

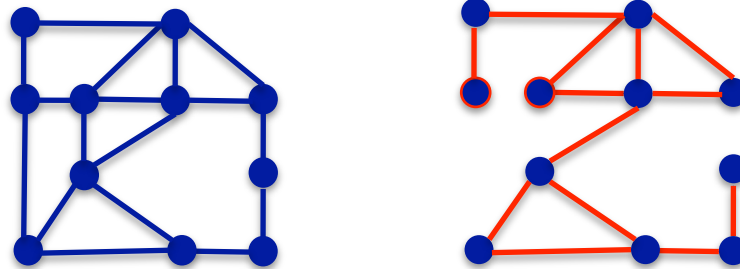


AND EVEN MORE COMMUNITIES



CHORDAL GRAPHS

- A chordal graph has no chordless cycles with more than 3 nodes



- various problems that are NP-hard on general graphs can be solved in polynomial time on chordal graphs
- turns up in various contexts



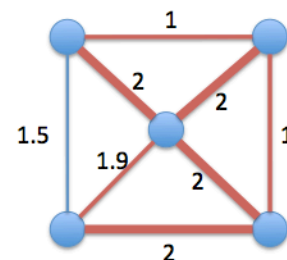
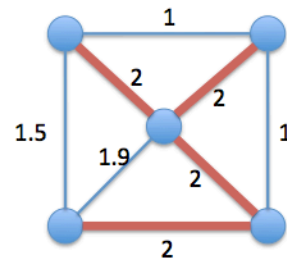
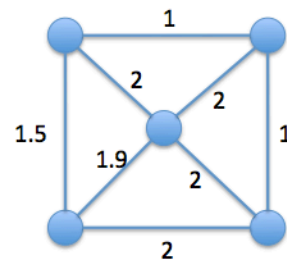
MAXIMAL CHORDAL SUBGRAPHS

- Existing algorithms (e.g. Dearing-Shier-Warner 1988) are based on Rose-Tarjan-Lueker theorem (1976) stating that C is a maximal chordal subgraph of G if and only if
 - C has a perfect elimination ordering (v_1, v_2, \dots, v_n) , and
 - for all edges $(v_i, v_j) \in G \setminus C$ with $i < j$, the subgraph of $C \cup (v_i, v_j)$ induced by the vertices v_i, v_{i+1}, \dots, v_n is not chordal.
- Algorithm:
 - Select vertex v
 - Add v , and all edges incident on v that maintain chordality, to chordal subgraph
 - Continue until all vertices are selected



A DIFFERENT APPROACH

- Given a connected chordal graph C and an edge (u,v) , adding (u,v) to C maintains chordality if and only if there is no path from u to v that does not go through a shared neighbor. (Markenzon-Vernet-Araujo theorem (2008).)



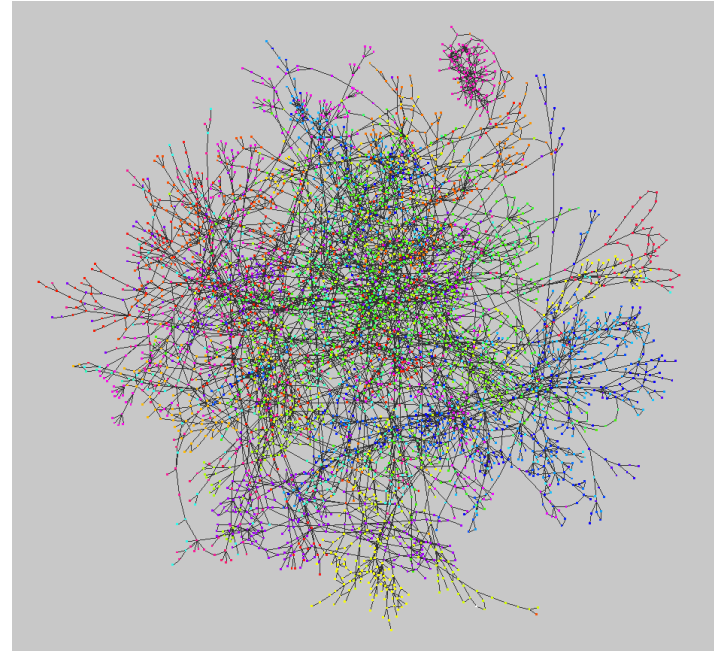
Input	Initialization	Augmentation	
	Num edges	Edges added	Num iters
brca	48,802	75,116	5
ctrl	48,802	118,933	5
cnr	32,556	357,917	4
gt	1,123,916	929,337	9
RGG18	262,091	181,960	7
RGG19	524,224	365,993	7
RGG20	1,048,477	739,008	7
RGG21	2,097,024	1,483,146	7
RGG22	4,194,120	2,990,974	7
RMAT21	2,067,408	261	1
RMAT22	4,128,561	262	1
RMAT23	8,243,779	257	1
RMAT24	16,460,670	328	1
RMAT25	32,864,788	363	1
RMAT26	65,617,354	388	1

A new augmentation based algorithm for extracting maximal chordal subgraphs
Sanjukta Bhowmick, Tzu-Yi Chen, Mahantesh Halappanavar.
JPDC: v76, pg 145-157, February 2015.



BACK TO COMMUNITIES

- Ground truth?
- Stability?
 - Input: graph $G=(V,E)$
 - compute clustering C of G
 - repeat:
 - perturb G to form $G'=(V,E')$
 - compute clustering C' of G'
 - measure distance between C' and C

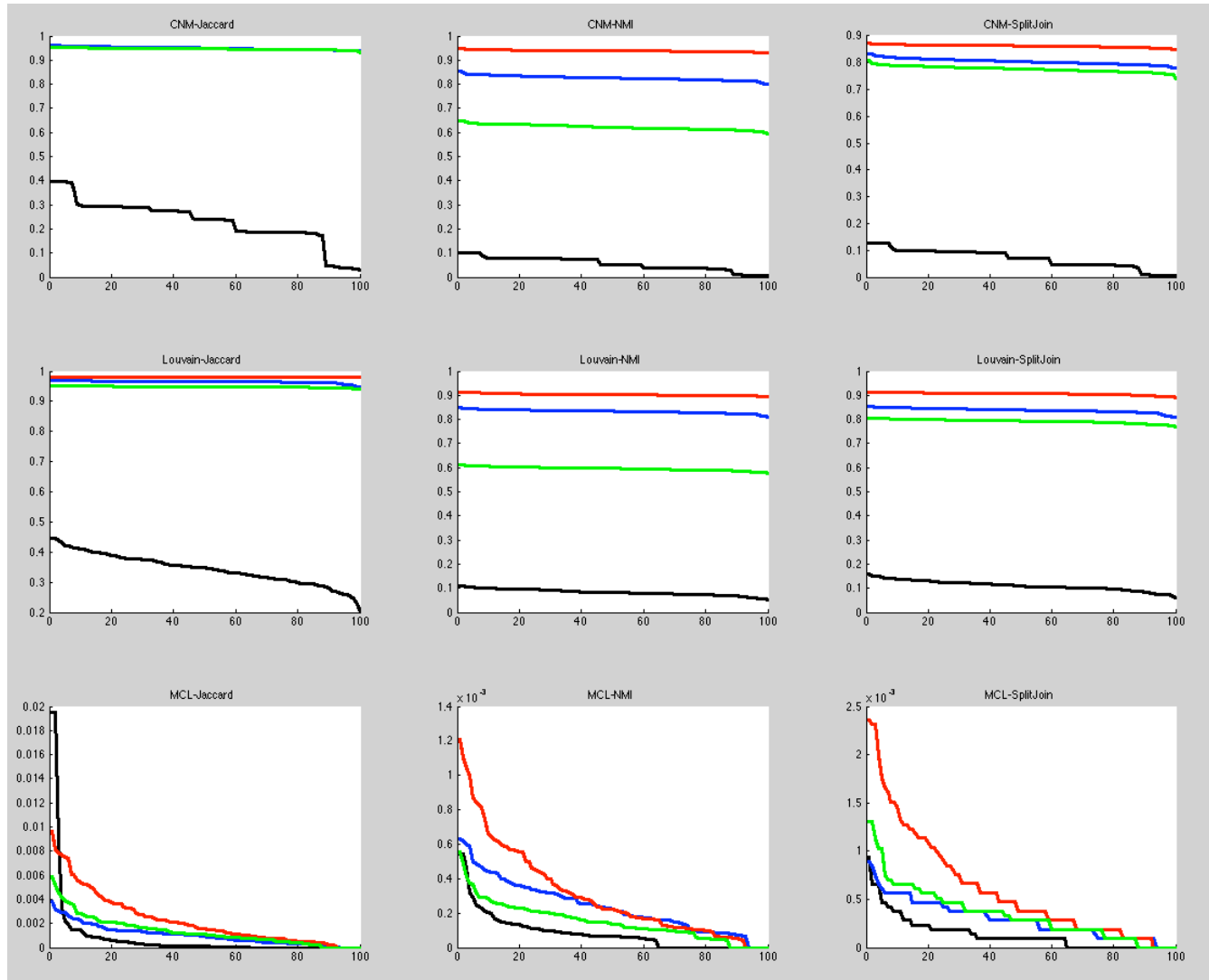


Jaccard

NMI

Split-Join

CNM



Louvain

MCL

Evaluating the stability of communities found by clustering algorithms.
Tzu-Yi Chen and Evan Fields (Pomona '13). CompleNet '13.



WHY MODULARITY?



- Assumptions?
- Alternatives?
- Impact on heuristics of assuming something else?



THE PIPELINE – SUMMER 2015



- SURP: Summer Undergraduate Research Program
- HAP: High Achievement Program
- PAYS: Pomona Academy for Youth Success



“David Bindel called last night [and] said that Jim is very proud of my getting the NSF CAREER award. Evidently he's mentioned it at least 4 times in various contexts/at various places. You know, I think the fact that Jim is proud means almost as much to me as getting the award in the first place. Almost. :)”

Diary entry, 6/22/2005

