

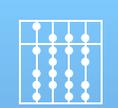
***Dichte-basiertes Clustering zur
dynamischen und abstrakten
Darstellung von großen Netzwerken***

Klaus Holzapfel

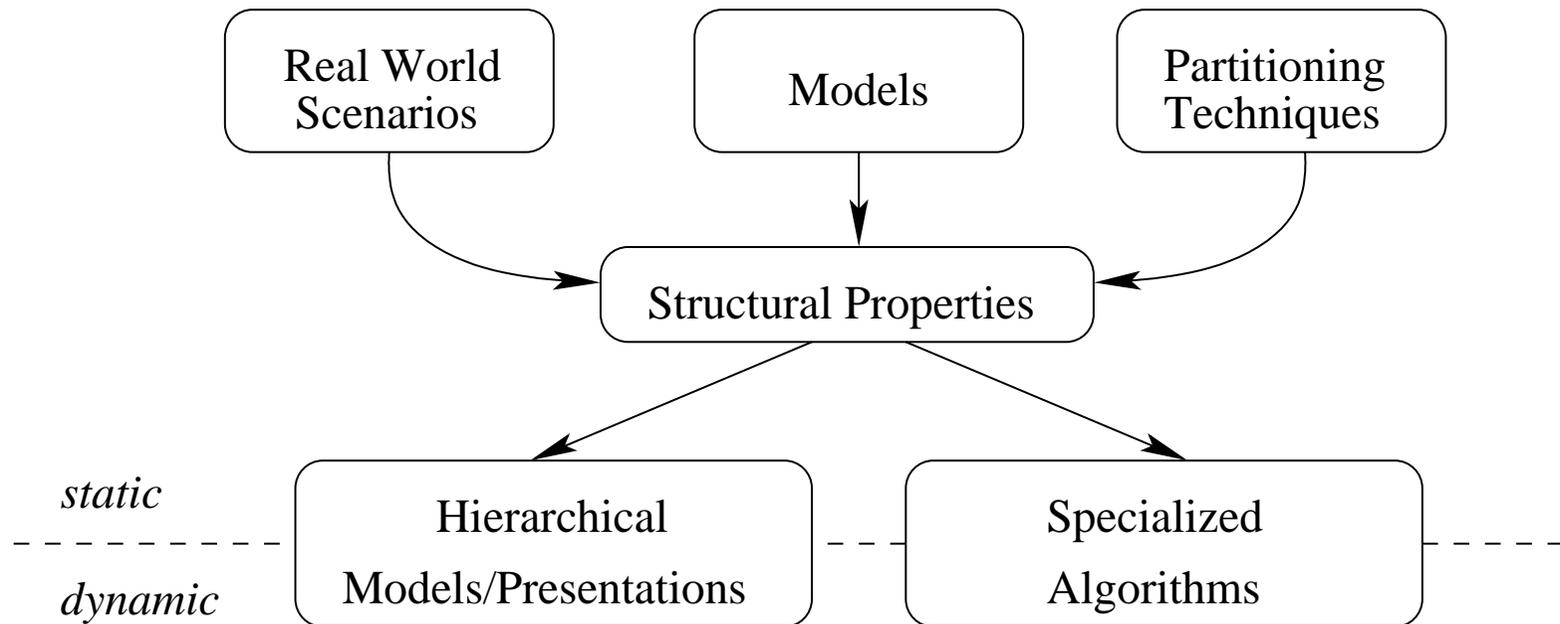
Lehrstuhl für Effiziente Algorithmen

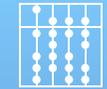
Fakultät für Informatik an der Technische Universität München

Workshop "Webgraphen" – 18. Okt 2002 (Tübingen)



Projektübersicht und Ausblick auf das Thema





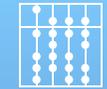
Beispiele:

- ▶ Internet
- ▶ Hyperlinkstruktur des WWW
- ▶ Telefon- und Stromnetzwerke
- ▶ Metabolische Netze

Eigenschaften:

- ▶ Small-World Graphen:
 - dünne Graphen ($\bar{d} < 20$)
 - geringer Durchmesser ($\sim \log(|V|)$)
 - Clustertendenz
- ▶ Power-Law Graphen:

$$\#\text{deg}_i(G) \sim i^{-\alpha} \quad \alpha \in [2, 4]$$



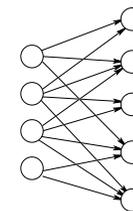
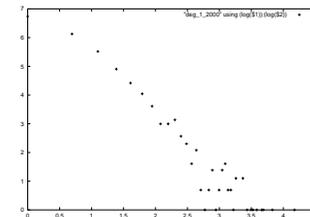
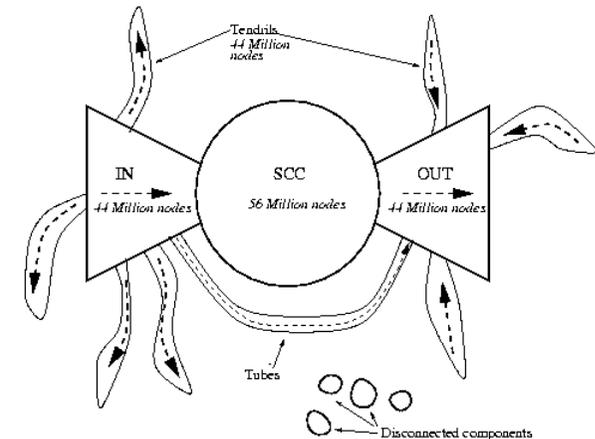
Probleme:

- ▶ riesige Datenmengen ($|V| \approx 10^9$)
- ▶ keine globale Struktur
- ▶ unklare Verbindungsregeln
- ▶ herkömmliche Algorithmen ungünstig

Lösungsansätze:

- ▶ Abstraktion:
 - VLSI Design
 - Transportprobleme
 - Gruppierungsprobleme
- ▶ Hierarchisierung
- ▶ neue Modelle mit Algorithmen

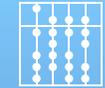
- ▶ kleiner Durchschnittsgrad
 - ▶ Clustertendenz (Communities)
 - ▶ gerichtet, lange Pfade
 - ▶ kein Small-World (klassisch)
-
- ▶ Power-Law Verteilung ($\alpha \approx 2.3$)
-
- ▶ Hubs and Authorities
 - ▶ Cores G_{ij} in Communities



Dynamische Hierarchisierung des WWW mittels Repräsentation durch Web-Communities.

Hierarchisches dichte-basiertes Modell von zeitlich variablen Small-World Graphen mit nach Power-Law verteilten Graden.

Anwendung einer solchen Hierarchie



WWW-Suche: Fibonacci

Netscape: Web search: Fibonacci

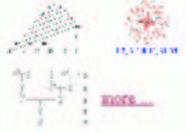
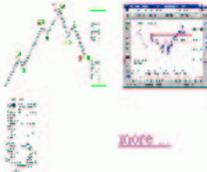
File Edit View Go Communicator Help

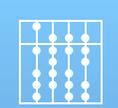
Location: file:/home/holzapfe/Working/Vortraege/SPP_1126_Jahrestreffen_02/Fibonacci/index.html

Back Forward Reload Home Search Netscape Print Security Shop Stop

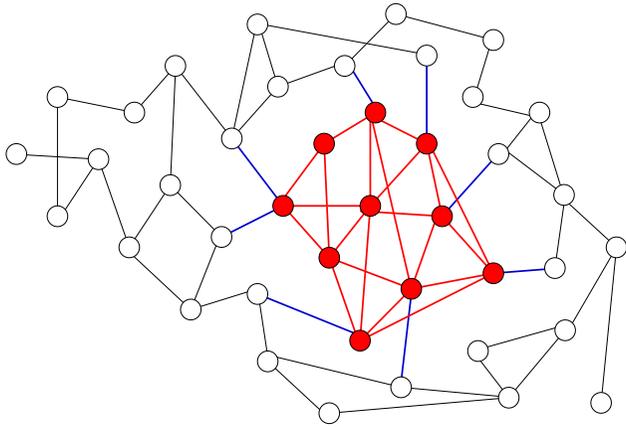
Lehrstuhl Google English/German Telefonliste RoomFinder Speiseplan Google Glossary

Search Phrase: Fibonacci

fibonacci: leonardo, pisa			fibonacci: number		
images	direct search results:	community: biography	images	direct search results:	community: mathematics, science
 <p>more...</p>	<ul style="list-style-type: none"> • Biography of Leonardo Fibonacci • Who was Fibonacci? • Fibonacci Leonardo da Pisa (ca. 1170-ca. 1240)... • more... 	<ul style="list-style-type: none"> • Biography Who's Who... • Biography • Biographical Dictionary Search Page • more... 	 <p>more...</p>	<ul style="list-style-type: none"> • What is a Fibonacci Number • Fibonacci Numbers... • FIBONACCI Numbers - ... • more... 	<ul style="list-style-type: none"> • Algebra • Fibonacci Heaps • Number Theory • more...
fibonacci: economics, stock markets			fibonacci: association		
images	direct search results:	community: economic, stock markets	images	direct search results:	community: associations, mathematical
 <p>more...</p>	<ul style="list-style-type: none"> • The Fibonacci Series and the Market • Stock Market analysis with the Fibonacci... • Do Fibonacci Numbers Predict Stock Prices? • more... 	<ul style="list-style-type: none"> • Monetary Economics and History... • Economics Data and Information Links... • Glossary • more... 	 <p>more...</p>	<ul style="list-style-type: none"> • The Fibonacci Association • Fibonacci Quarterly Home Page • The Tenth International Fibonacci Conference • more... 	<ul style="list-style-type: none"> • Societies and Associations • Math on the Web Societies... • Mathematical Associations and Societies • more...



bisherige Forschungsergebnisse



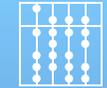
- ▶ viele interne Kanten (Dichte)
- ▶ wenige externe Kanten (Schnitt)
- ▶ viele kurze Pfade (Zusammenhang)

Problem: DENSE k -SUBGRAPH-PROBLEM

Input: Graph G , $k \in \mathbb{N}$

Output: Subgraph G' mit maximaler Anzahl von Kanten und genau k Knoten

- ▶ (variables) Entscheidungsproblem \mathcal{NP} -vollständig
- ▶ $\mathcal{O}(n^{\frac{1}{3}-\epsilon})$ -Approximation [Feige, Kortsarz, Peleg, 2001]



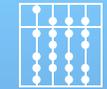
Problem: γ -DENSE SUBGRAPH PROBLEM (γ -DSP)

Input: Graph G , $k \in \mathbb{N}$

Output: Gibt es einen Subgraphen G' mit k Knoten und mindestens $\gamma(k)$ Kanten?

- $\gamma(k) = \binom{k}{2}$ γ -DSP = CLIQUE $\in \mathcal{NP}$ -c
- $\gamma(k) = 0$ γ -DSP $\in \mathcal{P}$

Wo liegt der Übergang?

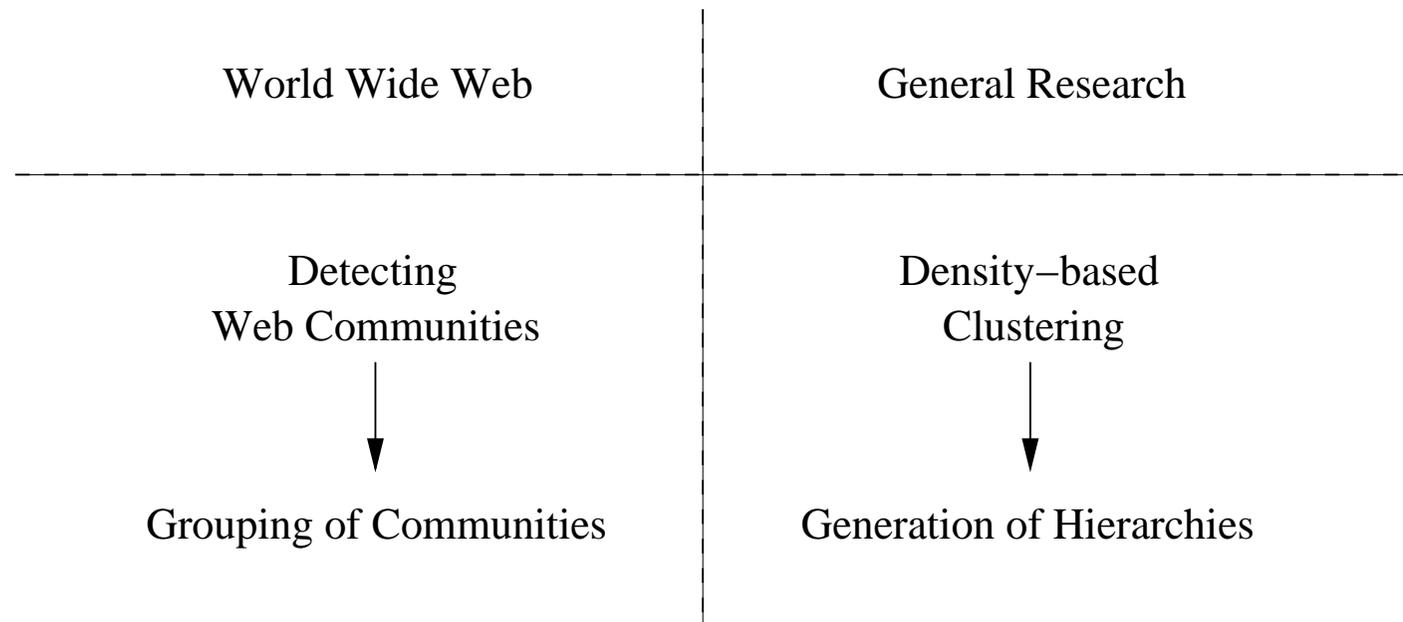


Ergebnisse – Übersicht

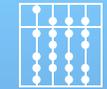
	\mathcal{P}	$\mathcal{NP-c}$
[Asahiro et.al. 2002]	$\gamma(k) = k$	$\gamma(k) = \Theta(k^{1+\epsilon})$
[Feige, Seltser 1997]		$\gamma(k) = k + k^\epsilon$
[H et.al. 2002]	$\gamma(k) = k + O(1)$	$\gamma(k) = k + \Omega(k^\epsilon)$

Forschungsideen

Was kommt als nächstes?



- *Trawling the Web for Emerging Cyber Communities*
[Kumar, Raghavan, Rajagopalan, Tomkins, 1999] WWW8
- *An approach to build a cyber-community hierarchy*
[Krishna Reddy, Kitsuregawa, 2002] Workshop on Web Analytics

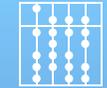


Studentenprojekte:

- ▶ Web-Crawler zum Sammeln von Daten
- ▶ Extrahieren von Web Communities

Forschungsarbeit:

- ▶ Begriffsklärung: Dichte in gerichteten Graphen
- ▶ Aufbau von Web-Hierarchien
 - Entwickeln von bottom-up Verfahren
 - Qualitätsaussagen (Approximationsgüte)



Integration in eine Hierarchie

- ▶ Bewertung von neuen Seiten
- ▶ Anpassen der Hierarchiebeschreibung (qualitativ und quantitativ)
- ▶ → “Passt die Hierarchie zum Datensatz?”

Restrukturierung der Hierarchie

- ▶ Gütekriterien für Hierarchien
- ▶ Nachführungen beim Auflösen bzw. Einfügen von Abstraktionsstufen
- ▶ Qualitätsaussagen nach mehreren Modifikationen