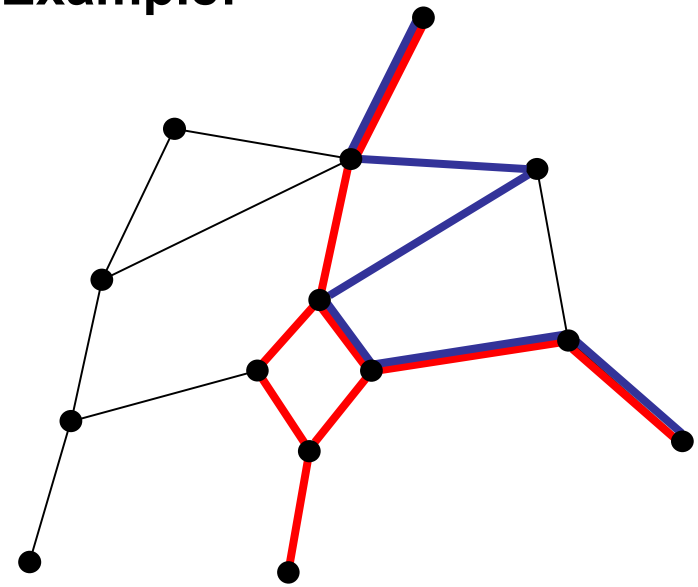
Closed Pattern Mining in Strongly Accessible Set Systems

Mario Boley, Tamás Horváth, Axel Poigné, Stefan Wrobel
Fraunhofer Institute IAIS & University of Bonn, Germany

given finite set E,
membership oracle $M_{\mathcal{F}}$: $2^E \rightarrow \{0,1\}$ (defining $\mathcal{F} \subseteq 2^E$),
closure operator ρ: $\mathcal{F} \rightarrow \mathcal{F}$,
list the family of ρ-closed sets $\rho(\mathcal{F}) = \{C \in \mathcal{F}: C = \rho(C)\}$

Example:



street network: graph G

transaction database $\mathcal{D}$: subsets of E(G)

- E: edge set E(G)
- $\mathcal{F}$: {X⊆E(G) : X is freq. and *connected*}
- ρ: $X \mapsto$ connected component of X in $\bigcap \{T \in \mathcal{D}: T \supseteq X\}$

main result:
problem can be solved efficiently with a simple DFS algorithm for a general class of set systems (which for instance do not satisfy anti-monotonicity)

set system $(E,\mathcal{F})$ is **strongly accessible** if $\emptyset \in \mathcal{F}$ and for every $X,Y \in \mathcal{F}$ satisfying $X \subset Y$, there exists an $e \in Y \setminus X$ such that $Y \setminus \{e\} \in \mathcal{F}$.

⇨ there is a sequence $X = X_0, X_1, \ldots, X_k = Y$ s.t. $|X_i \setminus X_{i-1}| = 1$ for $i = 1,\ldots,k$

formally the following statement holds:

Thm: For any finite strongly accessible set system $(E,\mathcal{F})$

(i) given by a membership oracle M computable in time T_M,

(ii) for any closure operator $\rho: \mathcal{F} \to \mathcal{F}$, computable in time T_ρ

the family of closed sets can be listed with delay $(|E|^2 T_M T_\rho)$.

Details at the poster!

universitätbonn
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT
Fraunhofer IAIS Institut Intelligente Analyse- und Informationssysteme