Sociálne siete a fuzzy formálna konceptová analýza

Stanislav Krajči[1], Jana Krajčiová[2]

[1]Ústav informatiky, Prírodovedecká fakulta, UPJŠ Košice
stanislav.krajci@upjs.sk
[2]Gymnázium, Alejová 1, Košice
jana.krajciova@pobox.sk

Abstrakt Jedným z dôležitých faktorov efektívneho team-buildingu je rešpektovanie medziľudských vzťahov. Personálny manažér zverenej sociálnej siete by preto mal poznať jej štruktúru aj z tohto hľadiska. V našom príspevku sa zaoberáme špeciálnym prípadom sociálnej siete – istou triedou osemročného gymnázia, na ktorej sme urobili takýto experiment: Každý žiak najprv charakterizoval svoj vzťah ku každému spolužiakovi hodnotou z istého rozsahu. Na takto získané relačné dáta sme aplikovali data-miningovú zhlukovaciu metódu formálnej konceptovej analýzy – jednostranne fuzzy konceptové zväzy včítane príslušne modifikovaného Rice-ovho a Siff-ovho algoritmu. Pomocou neho získané zhluky – skupiny žiakov vnímaných svojimi spolužiakmi rovnakým spôsobom – sme kvalifikovane interpretovali ako užšie či širšie priateľské zväzky medzi nimi. V závere diskutujeme možnosť aplikácie tohto postupu na iné sociálne siete, hlavne vo firemnej praxi. Takto získaná znalosť môže pomôcť personálnemu manažérovi firmy zložiť jednotlivé jej tímy tak, aby prípadné zlé medziľudské vzťahy ich členov zbytočne nebránili jej rozvoju.

1 Úvod

Sociálna sieť môže byť charakterizovaná ako systém vzájomných vzťahov medzi ľuďmi istej skupiny. Takáto sieť je často vyjadrená vo forme orientovaného grafu, ktorého vrcholy tvoria účastníci siete. V najjednoduchšom prípade prítomnosť (orientovanej) hrany znamená iba existenciu vzťahu osoby reprezentovanej začiatočným vrcholom tejto hrany k osobe reprezentovanej jej koncovým vrcholom. (Poznamenajme, že takýto vzťah nemusí byť vzájomný, orientovanosť hrán je preto pochopiteľná.) Takýto graf môže byť reprezentovaný klasickou reláciou: Ak B je množina zúčastnených osôb, tak naša relácia je istá podmnožina množiny $B \times B$. Ak je táto relácia znázornená vo forme tabuľky, množina B znamená jednak jej riadky (ako objekty) a jednak stĺpce (ako atribúty). Každá hodnota v tejto tabuľke potom znamená ohodnotenie vzťahu osoby zo stĺpca k osobe z riadku.

Toto je však ideálna situácia na aplikovanie nejakej data-miningovej metódy na získanie prípadných skrytých informácií z tejto tabuľky. V tomto článku sa

* S podporou grantu 1/3129/06 Slovenskej grantovej agentúry VEGA.

© Václav Snášel (Ed.): Znalosti 2008, pp. 101–111, ISBN 978-80-227-2827-0.
FIIT STU Bratislava, Ústav informatiky a softvérového inžinierstva, 2008.

sústredíme iba na jednu z nich, na *formálnu konceptovú analýzu* (čím, samozrejme, nevylučujeme použitie iných prístupov). Ako výsledok tak budeme očakávať koncepty, t. j. zhluky osôb, ktoré sú považované inými osobami za podobné.

Klasická podoba formálnej konceptovej analýzy pracuje s dvoma hodnotami – áno, resp. nie. Existuje však mnoho situácií, keď tieto dve hodnoty nedokážu skúmané vzťahy plne popísať. V takom prípade potrebujeme na ich charakterizáciu viac hodnôt, tie budú hodnotami (farbami, váhami) hrán grafu popisujúceho sociálnu sieť. Hovoríme tu potom o *fuzzy relácii*, ktorá nie je ničím iným ako funkciou z množiny $B \times B$ do predpísanej množiny hodnôt (ktorá je často usporiadaná, ba dokonca lineárne). Ak tu chceme použiť formálnu konceptovú analýzu, musíme sa uchýliť k niektorej jej viachodnotovej verzii. V našom experimente používame tzv. *jednostranne fuzzy konceptový zväz*.

2 Experiment

Náš experiment bol vykonaný na gymnáziu, kde ako učiteľka pracuje i spoluautorka tohto článku. Vybrali sme triedu *Sekunda*, ktorú dobre pozná, lebo bola skoro rok a pol jej triednou učiteľkou. Túto triedu navštevuje 34 okolo 12-ročných študentov, z toho 11 dievčat a 23 chlapcov.

Každý študent dostal za úlohu vyjadriť svoj vzťah ku každému spolužiakovi výberom jednej z nasledujúcich siedmich hodnôt (poznamenajme, že tento počet zodpovedá známej Lickertovej stupnici):

- 3 – je to môj výborný priateľ
- 2 – je to môj kamarát
- 1 – je mi sympatický
- 0 – nie je mi ani sympatický, ale ani nesympatický
- -1 – nie je mi sympatický
- -2 – nemám ho rád
- -3 – neznášam ho

Všimnime si, že hodnoty sú stupňované, väčšia znamená lepší vzťah. Hodnoty boli vybraté v súlade s bežným chápaním celých čísel, teda tak, aby kladné vyjadrovali pozitívny vzťah, záporné negatívny a nula, prirodzene, neutrálny. Treba zdôrazniť, že tu nedorozumenie zo strany žiakov nehrozilo – v tomto veku už veľmi dobre rozumejú významu záporných čísel, navyše je táto trieda explicitne zameraná na matematiku. Žiakov sme motivovali prísľubom, že sa dozvedia (aspoň čiastkové) výsledky experimentu. Možno aj preto k ankete pristúpili veľmi zodpovedne, vypĺňali ju v dobrej atmosfére, bez časového stresu i bez iných rušivých elementov.

Výsledky sú v nasledujúcej tabuľke. Stĺpec každej hodnoty znamená žiaka, ktorý hodnotil, jej riadok zodpovedá žiakovi, ktorý bol hodnotený. Treba povedať, že hodnoty na diagonále (t. j. samohodnotenia študentov) boli upravené na jednotnú, prirodzene najvyššiu, hodnotu (väčšina z nich ju naozaj použila, no niektorí túto položku vôbec nevyplnili a pár ich použilo nižšie (dokonca záporné) hodnoty). Všetky ostatné hodnoty však ostali, samozrejme, nezmenené.

Tabuľka 1 Ohodnotenia vzťahov medzi žiakmi

	name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	Ch01	3	−2	1	1	3	0	−1	0	3	0	−2	2	−2	1	1	0	0	0	3	2	2	0	−1	−1	0	−1	0	0	0	0	3	0	0	0
2	Ch02	1	3	3	0	0	0	3	3	0	1	3	0	3	3	2	−1	2	1	−1	0	1	0	1	0	−2	2	−1	0	2	0	1	2	3	2
3	Ch03	0	3	3	2	−1	−2	3	2	1	3	3	1	2	3	2	−2	1	0	0	1	1	0	0	0	1	0	0	1	2	0	2	0	2	1
4	Ch04	−1	−1	0	3	−1	−2	−2	0	0	0	−1	2	−1	−3	1	−2	−2	0	−2	1	−1	0	−2	−1	3	−1	0	2	0	0	0	0	0	−2
5	Ch05	3	0	2	−1	3	0	−1	0	2	0	1	2	0	2	2	0	1	1	3	2	2	0	0	0	0	0	1	0	0	0	3	0	0	1
6	D01	−1	−1	−1	−1	−1	3	−3	0	−1	0	0	0	−2	−2	1	3	1	2	−1	0	−1	2	2	2	0	2	1	0	−2	2	0	1	0	2
7	Ch06	0	3	3	−1	−2	0	3	3	−3	0	3	0	3	3	2	0	1	1	−2	0	0	0	1	0	0	2	1	1	3	0	−1	2	3	1
8	Ch07	1	2	3	1	1	0	2	3	1	1	3	1	2	3	2	0	2	1	1	0	1	0	1	0	2	1	1	1	1	0	2	0	3	2
9	Ch08	3	1	2	2	3	−1	−1	−1	3	1	1	2	−3	−3	2	−1	2	1	3	1	3	0	1	0	−3	1	1	0	2	0	3	0	−2	1
10	Ch09	0	0	3	1	0	−2	0	2	1	3	0	1	−2	−1	1	−1	1	0	1	1	0	0	0	−1	0	0	1	1	1	0	0	−1	1	1
11	Ch10	−1	3	3	−2	−1	0	3	3	−3	0	3	0	3	3	3	−2	2	1	−2	0	0	0	1	0	−3	2	−1	0	2	0	−2	2	3	2
12	Ch11	1	0	1	2	2	0	−2	0	0	0	0	3	0	−2	1	0	−2	1	−1	3	0	0	−1	−2	1	−1	1	1	1	0	2	0	0	0
13	Ch12	−1	3	3	0	−1	−2	3	3	−3	−1	3	1	3	3	2	−2	2	1	0	1	1	−1	−1	−2	0	0	−2	1	2	0	−1	−2	3	2
14	Ch13	−1	3	2	−1	−1	−3	1	3	−1	−1	3	0	2	3	2	−2	3	0	−3	0	0	0	0	−2	0	0	0	0	0	0	0	−2	2	3
15	Ch14	−2	2	3	−1	−1	0	1	2	1	0	3	0	2	2	3	−1	3	1	0	0	1	0	0	−1	−1	0	1	1	3	0	0	0	2	3
16	D02	−1	0	−2	0	−1	3	0	−1	0	−1	0	0	−2	−3	1	3	2	2	−1	0	0	2	2	2	0	2	1	0	−1	2	0	2	0	2
17	Ch15	−3	−1	0	−3	−3	−1	−1	0	1	−2	0	0	−3	−3	3	−2	3	1	0	−3	0	−1	−1	0	−3	0	0	−1	−2	−1	−3	0	0	3
18	D03	−1	1	−1	0	−1	2	1	−1	0	0	3	0	1	−3	3	2	3	3	−1	0	0	0	2	2	0	2	3	0	1	1	2	3	−1	2
19	Ch16	3	0	0	−3	3	0	−1	−1	3	1	0	0	−1	−3	2	0	2	1	3	1	2	0	0	0	−3	0	1	1	2	0	3	1	−1	1
20	Ch17	0	−2	−1	1	1	0	−3	−1	−3	−2	−2	3	−3	−3	−3	0	−3	0	−2	3	−2	0	−1	−2	1	−1	1	−1	0	0	0	0	−1	−2
21	Ch18	3	1	1	1	3	−1	1	2	3	1	1	2	1	1	2	−1	1	0	3	2	3	0	0	0	0	0	1	0	2	0	3	1	1	0
22	D04	−1	0	−1	0	−1	0	0	−1	−2	0	−1	0	−3	−3	2	0	−1	1	−1	0	−2	3	2	1	0	2	1	0	0	1	0	1	−2	2
23	D05	−1	−1	−1	0	−1	1	0	−1	−3	0	0	0	−2	−3	2	1	2	2	−1	0	−1	2	3	1	0	3	2	0	−2	1	0	1	−1	2
24	D06	−1	0	0	0	−1	3	−3	0	0	0	0	0	0	−2	2	3	3	2	−1	0	0	2	2	3	0	2	2	0	−1	3	0	2	0	3
25	Ch19	−1	−3	−3	3	−1	−3	−3	−1	−3	−1	−3	1	−2	−3	−3	−3	−3	0	−3	1	−2	0	−3	−2	3	−3	−2	2	−3	−1	−1	−1	−1	−3
26	D07	−1	1	−1	0	−1	1	2	0	−3	0	3	0	1	−3	3	1	3	2	−1	0	0	2	3	1	0	3	2	0	2	1	0	1	0	2
27	D08	−2	−3	−1	0	−1	1	1	−2	0	0	−3	0	−2	−3	3	1	3	3	−1	0	0	0	2	2	0	2	3	0	1	1	2	3	−1	−1
28	Ch20	−1	−3	−2	3	0	−2	−2	−1	−3	0	−3	1	−2	−3	−1	−2	−2	0	−2	1	−2	0	−1	0	3	−1	0	3	1	0	0	1	0	0
29	Ch21	0	2	3	2	0	0	3	2	1	1	3	1	1	1	3	−1	3	1	−1	1	1	0	1	0	1	1	2	1	3	0	0	1	2	2
30	D09	−1	−1	−2	0	−1	3	−3	−1	0	0	0	0	−1	−2	2	2	1	2	−1	0	0	2	2	3	0	2	1	0	−1	3	0	1	0	2
31	Ch22	3	−1	2	1	3	0	0	0	3	0	−1	3	−2	−1	2	0	2	1	3	2	3	0	0	0	1	0	1	1	2	0	3	1	0	0
32	D10	−1	2	1	0	−1	2	2	1	0	0	3	0	−3	−2	2	1	3	2	−1	0	0	2	2	2	0	2	2	0	2	1	1	3	0	2
33	Ch23	1	3	3	1	−1	0	3	3	−3	1	3	1	1	3	2	0	3	1	−1	2	0	0	1	0	0	1	2	1	2	0	1	0	3	1
34	D11	0	0	1	−2	0	2	−2	1	0	0	2	0	1	0	3	2	3	1	−1	0	0	2	2	2	−3	2	1	1	1	1	0	1	1	3

3 Jednostranne fuzzy konceptový zväz

Naša tabuľka dát je jednoducho objektovo-atribútový model: Riadky sú žiaci, čiže objekty nášho výskumu, stĺpce zodpovedajú hodnotiacim spolužiakom a môžeme ich chápať ako atribúty. Naše (ako sme už naznačili, psychologicky užitočné) hodnoty z množiny $\{-3,-2,-1,0,1,2,3\}$ možno ľahko transformovať (pomocou funkcie $x \mapsto \frac{x+3}{6}$) na hodnoty z množiny $\{0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1\}$, čo je podmnožina klasického fuzzy intervalu $[0,1]$. Hodnoty v každom riadku zodpovedajú jednému objektu a môžu byť chápané ako jemu zodpovedajúca funkcia z množiny všetkých atribútov do $[0,1]$. Budeme teda pracovať s klasickými (tzv. *crisp*) podmnožinami objektov a fuzzy podmnožinami atribútov.

V článku [K] sme modifikovali klasickú binárnu formálnu konceptovú analýzu z [GW] na tzv. *jednostranne fuzzy konceptový zväz*, ktorý pracuje práve s klasickými podmnožinami objektov a fuzzy podmnožinami atribútov. Tento prístup je v podstate ekvivalentný tzv *škálovaniu* Gantera & Willeho ([GW]), môžeme ho však považovať za jeho rýchlejšiu a efektívnejšiu verziu. (Poznamenajme, že podobné prístupy sa vyskytujú i v prácach Ben Yahiu a Jaouu v [BJ] a Bělohlávka a iných v [BSZ].)

Pripomeňme základné pojmy z teórie jednostranne fuzzy konceptových zväzov, ako boli navrhnuté v našom článku [K]:

Nech A (ako <u>a</u>tribúty) a B (ako o<u>b</u>jekty) sú neprázdne (obvykle konečné) množiny a nech R je fuzzy relácia na ich karteziánskom súčine, t. j. $R : A \times B \to [0,1]$. Túto reláciu potom môžeme chápať ako tabuľku s riadkami a stĺpcami zodpovedajúcimi objektom, resp. atribútom, hodnota $R(a,b)$ vyjadruje stupeň, v akom má objekt b atribút a.

Definujme zobrazenie $\uparrow : \mathscr{P}(B) \to [0,1]^A$, ktoré priradí každej množine objektov X fuzzy množinu atribútov $\uparrow(X)$, pričom hodnota v bode $a \in A$ bude

$$\uparrow(X)(a) = \inf\{R(a,b) : b \in X\}.$$

Táto funkcia teda priraďuje každému atribútu najväčšiu možnú hodnotu takú, aby každý objekt z X mal tento atribút aspoň v takomto stupni.

Symetricky definujme zobrazenie $\downarrow : [0,1]^A \to \mathscr{P}(B)$, ktoré priradí každej funkcii $f : A \to [0,1]$ množinu

$$\downarrow(f) = \{b \in B : (\forall a \in A) R(a,b) \geq f(a)\},$$

čiže všetky objekty majúce všetky atribúty v aspoň takom stupni, ako predpisuje funkcia f.

Ľahko vidieť, že tieto zobrazenia majú nasledujúce vlastnosti:

- Ak $X_1 \subseteq X_2$, tak $\uparrow(X_1) \geq \uparrow(X_2)$.
- Ak $f_1 \leq f_2$, tak $\downarrow(f_1) \supseteq \downarrow(f_2)$.
- $X \subseteq \downarrow(\uparrow(X))$.
- $f \leq \uparrow(\downarrow(f))$.

Pod $f_1 \leq f_2$ pritom rozumieme, že pre všetky prvky x z definičného oboru (spoločného pre f_1 i f_2) platí $f_1(x) \leq f_2(x)$.

Tieto vlastnosti sú ekvivalentné tvrdeniu, že pre každé $X \subseteq B$ a $f \in [0,1]^A$ platí

$$f \leq \uparrow(X), \quad \text{akk} \quad X \subseteq \downarrow(f)$$

(alebo, ekvivalentne, že štvorica $\langle \uparrow, \downarrow, \subseteq, \leq \rangle$ je Galoisova konexia).

Poznamenajme, že v dvojhodnotovom prípade, t. j. keď je obor hodnôt relácie R (najviac) dvojprvková množina $\{0,1\}$, znamená $R(a,b) = 1$, že atribút a je atribútom objektu b, a opačne, $R(a,b) = 0$ znamená, že a nie je atribútom b. V tomto zmysle je tento prístup zovšeobecnením klasického Ganterovho-Willeho prístupu.

Teraz definujme zobrazenie $\mathrm{cl} : \mathscr{P}(B) \to \mathscr{P}(B)$ ako zloženie zobrazení $\uparrow$ a $\downarrow$: pre všetky $X \subseteq B$ položme $\mathrm{cl}(X) = \downarrow(\uparrow(X))$.

Opäť ľahko vidieť, že cl je operátor uzáveru, čiže že sú splnené nasledujúce tri podmienky:

- $X \subseteq \mathrm{cl}(X)$.
- Ak $X_1 \subseteq X_2$, tak $\mathrm{cl}(X_1) \subseteq \mathrm{cl}(X_2)$.
- $\mathrm{cl}(X) = \mathrm{cl}(\mathrm{cl}(X))$.

Ako v klasickom prípade, i tu hrajú dôležitú rolu tie množiny objektov X, pre ktoré $X = \mathrm{cl}(X)$ (pretože vtedy platí $f = \uparrow(X)$ práve vtedy, keď $X = \downarrow(f)$). Takúto dvojicu $\langle X, \uparrow(X) \rangle$ nazývame *jednostranne fuzzy koncept* (lebo $\uparrow(X)$ je fuzzy množina, kým množina X je klasická). Potom X nazývame *extent* takéhoto konceptu a zodpovedajúcu fuzzy množinu $\uparrow(X)$ jeho *intent.* Kvôli možnosti jednoznačného vzájomného určenia oboch súradníc stačí uvažovať len jednu z nich, často je to prvá: Množina $\{X \in \mathscr{P}(B) : X = \mathrm{cl}(X)\}$ usporiadaná inklúziou je potom (úplný) zväz s operáciami $X_1 \wedge X_2 = X_1 \cap X_2$ a $X_1 \vee X_2 = \mathrm{cl}(X_1 \cup X_2)$. Tento zväz nazveme *jednostranne fuzzy konceptový zväz.*

4 Rice-ov a Siff-ov algoritmus

Použitie formálnej konceptovej analýzy (či už klasickej alebo viachodnotovej) má (podobne ako mnoho iných zhlukovacích metód) jednu veľkú nevýhodu – spravidla priveľké množstvo nájdených zhlukov. Stalo sa to i v našom prípade – jednostranne fuzzy konceptový zväz vzniknutý z našich dát obsahoval viac než 25000 zhlukov. Uvažovať o každom koncepte jednotlivo je pri takomto množstve zrejme nemožné. Je preto potrebné hľadať doplnkové metódy na redukciu tejto škaredej vlastnosti. Niektoré z nich vynašiel tím Radima Bělohlávka (napríklad [BV]). Sú založené buď na využití nejakej dodatočnej informácie o objektoch alebo atribútoch (napr. ich usporiadanie alebo nejaká iná relácia), alebo na redukcii kardinality oboru hodnôt použitím tzv. *zdôrazňovačov pravdy.* V našom príklade však nemáme žiadnu dodatočnú informáciu tohto typu a nechceme ani umelo redukovať pomerne malý počet možných hodnôt (to by totiž znamenalo, že v ankete stačí respondentom poskytnúť menej alternatív). (Predsa sme však v tomto smere isté pokusy urobili, keď sme našich sedem hodnôt prirodzene zredukovali na dve, ako možné odpovede na otázku sympatie „+“ (3, 2 a 1) a „−“ (−3, −2, −1 a 0), avšak počet konceptov bol i naďalej priveľký, a to okolo 2000.)

Aby zhluky mohli byť pre používateľa aspoň v nejakom zmysle užitočné, musí ich byť buď dosť málo, alebo malá musí byť ich kardinalita. Na túto požiadavku existuje istá odpoveď, ktorá je na formálnej konceptovej analýze založená, avšak využíva i ďalšie, metrické vlastnosti. Je navrhnutá v našom článku [K] a teraz ju v krátkosti pripomenieme:

Definujme funkciu $\rho : \mathscr{P}(B) \times \mathscr{P}(B) \to \mathbb{R}$ tak, že pre $X_1, X_2 \subseteq B$ položme

$$\rho(X_1, X_2) = 1 - \frac{\sum_{a \in A} \min\{\uparrow(X_1)(a), \uparrow(X_2)(a)\}}{\sum_{a \in A} \max\{\uparrow(X_1)(a), \uparrow(X_2)(a)\}}.$$

Dá sa dokázať, že táto funkcia je na množine všetkých extentov $\{X \subseteq B : \mathrm{cl}(X) = X\}$ metrikou. Poznamenajme, že táto metrika je zovšeobecnením vzdialenostnej funkcie použitej Rice-om a Siff-om v [RS], a to

$$\rho'(X_1, X_2) = 1 - \frac{|X_1 \cap X_2|}{|X_1 \cup X_2|}.$$

Teraz vezmeme hierarchický zhlukovací algoritmus z [RS] a jednoducho nahradíme ich metriku ρ' našou ρ. Výsledný algoritmus môžeme vyjadriť nasledujúcim pseudokódom:

---

vstup A, B, R
$\mathcal{C} \leftarrow \mathcal{D} \leftarrow \{\mathrm{cl}(\{b\}) : b \in B\}$
kým ($|\mathcal{D}| > 1$) rob
{
  $m \leftarrow \min\{\rho(X_1, X_2) : X_1, X_2 \in \mathcal{D}, X_1 \neq X_2\}$
  $\mathcal{E} \leftarrow \{\langle X_1, X_2\rangle \in \mathcal{D} \times \mathcal{D} : \rho(X_1, X_2) = m\}$
  $\mathcal{V} \leftarrow \{X \in \mathcal{D} : (\exists Y \in \mathcal{D})\langle X, Y\rangle \in \mathcal{E}\}$
  $\mathcal{N} \leftarrow \{\mathrm{cl}(X_1 \cup X_2) : \langle X_1, X_2\rangle \in \mathcal{E}\}$
  $\mathcal{D} \leftarrow (\mathcal{D} \setminus \mathcal{V}) \cup \mathcal{N}$
  $\mathcal{C} \leftarrow \mathcal{C} \cup \mathcal{N}$
}
výstup $\mathcal{C}$

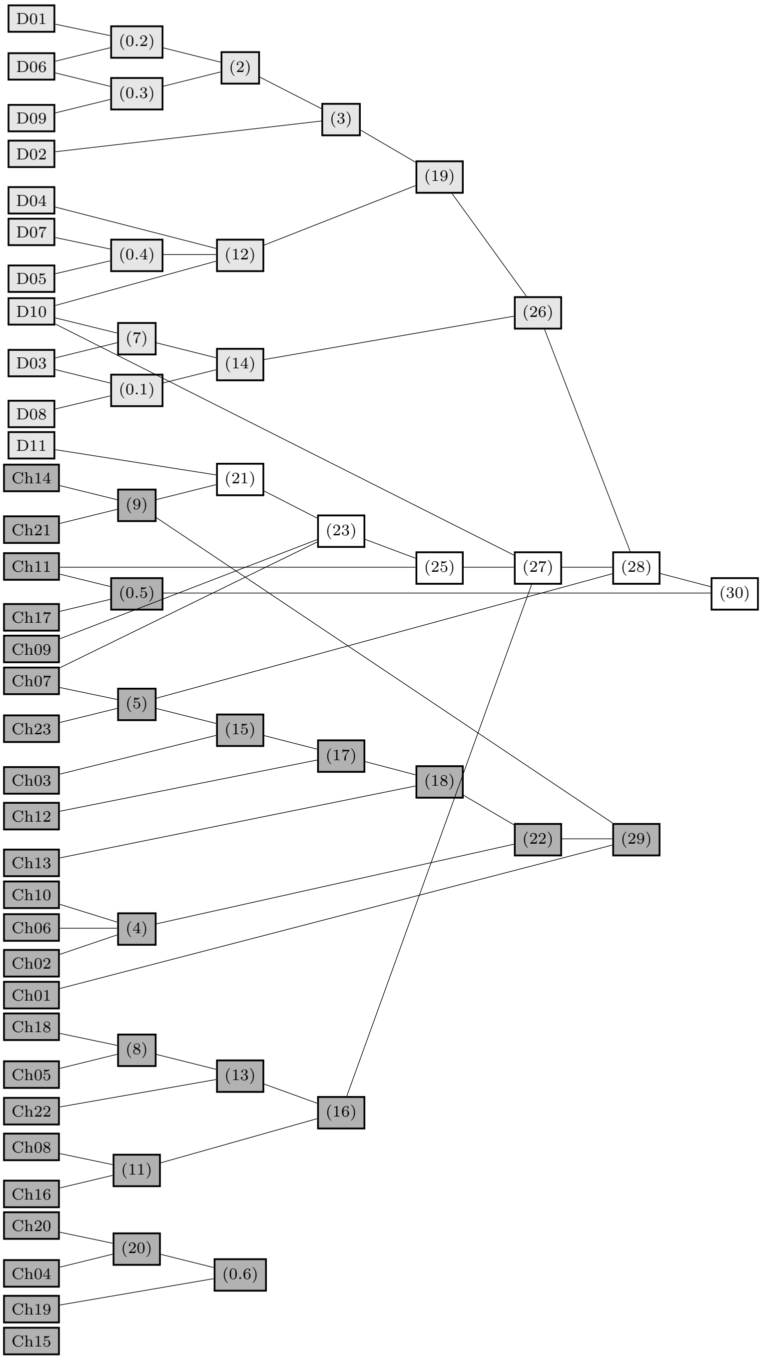
Premenná $\mathcal{D}$ je množina zhlukov-konceptov v aktuálnej iterácii, $\mathcal{C}$ je zjednotenie všetkých doterajších iterácií $\mathcal{D}$. Číslo m je najmenšia vzdialenosť dvojíc z $\mathcal{D}$. Množina $\mathcal{E}$ obsahuje „hrany“ – dvojice zhlukov z $\mathcal{D}$, ktorých vzdialenosť je práve m a premenná $\mathcal{V}$ obsahuje „vrcholy“ – konce takýchto hrán. Prvky množiny $\mathcal{N}$ sú nové zhluky. Nová iterácia $\mathcal{D}$ nahradzuje „staré“ zhluky z $\mathcal{V}$ „novými“ z $\mathcal{N}$.

Voľne povedané, v každom kroku spájame tie dva zhluky, ktoré sú najbližšie (v zmysle našej metriky) – z novej iterácie ich vyhodíme a namiesto nich do nej pridáme ich spojenie t. j. uzáver ich zjednotenia. Hoci tento algoritmus je v princípe exponenciálny, v prípade, keď v každej iterácii je najbližšie práve jedna dvojica (čo je zrejme pravda pre väčšinu reálnych dát), dostávame len okolo $2|B|$ zhlukov.

V našom prípade sme dostali nasledujúce zhluky. (Čísla v zátvorkách znamenajú kroky. Singletony (jednoprvkové množiny) sú vynechané pre ich zrejmú nezaujímavosť. Zhluky označené (0.*) sú ne-singletony, ktoré vznikli v 0. iterácii, sú uzávermi podčiarknutých objektov.)

(0.1) D03, <u>D08</u>
(0.2) D06, <u>D01</u>
(0.3) D06, <u>D09</u>
(0.4) D07, <u>D05</u>
(0.5) Ch11, <u>Ch17</u>
(0.6) Ch20, <u>Ch19</u>, Ch04
(1) rovnaký ako (0.3)
(2) D06, D01, D09
(3) D06, D01, D02, D09
(4) Ch10, Ch06, Ch02
(5) Ch23, Ch07
(6) rovnaký ako (4)
(7) D03, D10
(8) Ch18, Ch05
(9) Ch21, Ch14
(10) rovnaký ako (0.4)
(11) Ch08, Ch16
(12) D07, D10, D05, D04
(13) Ch18, Ch05, Ch22
(14) D03, D10, D08
(15) Ch23, Ch07, Ch03
(16) Ch08, Ch16, Ch18, Ch05, Ch22
(17) Ch12, Ch23, Ch07, Ch03
(18) Ch12, Ch23, Ch07, Ch03, Ch13
(19) D06, D07, D01, D10, D02, D05, D04, D09
(20) Ch20, Ch04
(21) D11, Ch14, Ch21
(22) Ch10, Ch12, Ch23, Ch07, Ch03, Ch13, Ch06, Ch02
(23) Ch09, D11, Ch07, Ch14, Ch21
(24) rovnaký ako (0.6)
(25) Ch09, D11, Ch11, Ch07, Ch14, Ch21
(26) D03, D06, D07, D01, D10, D02, D08, D05, D04, D09
(27) Ch09, D11, Ch08, Ch11, D10, Ch16, Ch07, Ch18, Ch05, Ch22, Ch14, Ch21
(28) Ch09, D03, D11, D06, D07, Ch08, Ch11, D01, D10, D02, D08, D05, Ch23, Ch16, Ch07, Ch18, Ch05, D04, Ch22, Ch14, Ch21, D09
(29) Ch10, Ch01, Ch12, Ch23, Ch07, Ch03, Ch13, Ch06, Ch14, Ch02, Ch21
(30) Ch09, D03, D11, D06, D07, Ch08, Ch11, D01, D10, D02, D08, D05, Ch23, Ch16, Ch07, Ch18, Ch05, Ch17, D04, Ch22, Ch14, Ch21, D09
(31) Ch09, D10, Ch10, Ch01, Ch12, Ch23, Ch07, Ch03, Ch13, Ch18, Ch05, Ch06, Ch14, Ch02, Ch15, Ch21
(32) všetci žiaci okrem Ch10, Ch01, Ch12, Ch13, Ch06 a Ch15
(33) všetci žiaci

Nasledujúci diagram zobrazuje tieto zhluky (s výnimkou posledných troch). Na ľavej strane sú všetky objekty, smerom doprava sú postupne spájané do väčších a väčších skupín. Tmavosivá farba je použitá na chlapcov a chlapčenské skupiny, svetlosivá znamená dievčatá a dievčenské skupiny. Zmiešané skupiny sú biele.



Obrázok 1 Väčšina zhlukov žiakov

5 Interpretácia

Naša interpretácia je založená na jednoduchom pozorovaní vyjadrenom slovenským príslovím „Vrana k vrane sadá“, teda že ľudia s podobnými charakteristikami sú priatelia (ak, pravdaže, majú na vytvorenie priateľstva dostatok času a príležitostí, čo je však v našom prípade zrejme splnené). Môžeme teda s veľkou dávkou istoty dedukovať, že zhluky, ktoré vznikli na základe vzájomných hodnotení spolužiakov, ktorí sa navzájom už dostatočne dlho poznajú, pozostávajú z osôb s dobrým vzájomným vzťahom, pričom čím je zhluk menší, tým je ich priateľstvo silnejšie. Pokúsme sa preto zhodnotiť naše zhluky z tohto hľadiska:

– Najviditeľnejšou črtou je dosť striktné delenie podľa pohlavia (existuje len 9 pohlavne zmiešaných zhlukov, 6 z nich je načrtnutých na diagrame). Toto pozorovanie korešponduje so všeobecne známym správaním 12-ročných detí. Jedinou výnimkou je D11, dievča s chlapčenskými záujmami typu karate.
– Veľmi zaujímavá skupina je (0.6), pozostávajúca z troch chlapcov, ktorí sú spolužiakmi v istom zmysle podceňovaní, pretože majú relatívne horší prospech a zároveň pochádzajú zo slabších sociálnych pomerov.
– Zhluk (31) (nie je na diagrame) je zaujímavý tým, že popri mnohých (hoci nie všetkých) chlapcoch obsahuje jediné dievča – D10, považované za krásku triedy, ktoré je navyše veľmi priateľské (ako vidieť, je spojivom dvoch menších dievčenských skupín).
– Ch15 je triedny exhibicionista. Sám často hovorí, že chce byť stredom pozornosti spolužiakov i učiteľov. To je pravdepodobne príčina, prečo nie je príliš obľúbený. Je iba v troch (na diagrame neznázornených) zhlukoch.
– Ako vo väčšine slovenských tried, aj tu žiaci sedia po dvojiciach. Tento rok im triedna učiteľka dala možnosť urobiť si tieto dvojice po svojom. Je zaujímavé, že sme v zhlukoch odhalili väčšinu z nich: D01 & D02, D06 & D09, D07 & D05, D03 & D08, D10 & D04, Ch14 & Ch21, Ch11 & Ch17, Ch07 & Ch23, Ch10 & Ch02, Ch04 & Ch19.
  Zaujímavé sú i dva páry Ch08 & Ch18 a Ch16 & Ch05. Pôvodné dvojice boli Ch08 & Ch16 and Ch18 & Ch05 (čiže podvojná výmena), ako to vidíme na diagrame. Chlapci však boli v tejto konfigurácii až príliš hluční, boli preto rozsadení.
  Zostávajúce páry D11 & Ch15 (vyššie spomenutí dvaja žiaci), Ch12 & Ch13, Ch22 & Ch01, Ch09 & Ch03 a Ch06 & Ch20 zostali neodhalené.

Samozrejme, vzniká prirodzená otázka, či existujú aj iné zaujímavé a podobne prirodzene interpretovateľné skupiny žiakov. Nepoznáme ich, nevylučujeme však, že existujú. Možno môžu byť odhalené nejakou inou zhlukovacou metódou.

6 Záver

V tomto článku sme popísali experiment s triedou žiakov istého košického gymnázia. Po vyhodnotení ich vzájomných vzťahov sme uvažovali ich sociálnu sieť ako viachodnotový objektovo-atribútový model a aplikovali jednostrannú fuzzifikáciu formálnej konceptovej analýzy. Počet vzniknutých konceptov – skupín žiakov – bol však priveľký, a preto v praxi nepoužiteľný. Redukovali sme ho preto, a to pomocou našej modifikácie Rice-ovho a Siff-ovho algoritmu, ktorá popri konceptovej analýze využíva aj isté metrické vlastnosti. Získali sme tak rozumný počet zhlukov a na naše milé prekvapenie sa ukázalo, že boli skutočne zaujímavé a ľahko interpretovateľné. Po tomto experimente sme presvedčení, že takáto aplikácia formálnej konceptovej analýzy môže pomôcť personálnemu manažérovi skupiny (v našom prípade triednemu učiteľovi) lepšie pochopiť štruktúru zverenej sociálnej siete.

Tento postup môže byť, pravdaže, použitý na ľubovoľnú relatívne uzavretú skupinu ľudí, napríklad na sociálnu sieť zamestnancov firmy. Ak sa navzájom dvaja pracovníci navzájom nepoznajú, ich vzťah bude charakterizovaný neutrálnou hodnotou. Pri ankete takéhoto osobného charakteru hrozí, že dospelí respondenti nebudú dostatočne úprimní, na získanie potrebných dát preto treba voliť pravdepodobne sofistikovanejšie metódy. To je už však skôr úloha psychológov.

Zaujímavou modifikáciou tohto experimentu by bola transpozícia matice vstupných dát (pripomeňme, že nemusí byť úplne symetrická), teda výmena hodnotiacich a hodnotených. Pri vytváraní zhlukov by potom nehrala úlohu podobnosť členov sociálnej siete z hľadiska ich kolegov, ale podobnosť ich vzťahov k ich spolupracovníkom. Tieto vzťahy sú síce obvykle v silnej korelácii, ale predsa sa môžu jemne líšiť. Takýto experiment sme však nezrealizovali.

Otvorenou ostáva otázka hodnotenia akejsi „pevnosti“ odhalených priateľstiev. Prvou naivnou aproximáciou je mohutnosť príslušného (pravdaže, nesingletonového) zhluku – čím menšia skupina, tým silnejšie priateľstvo –, do úvahy však prichádza i použitie vyššie spomínanej metriky.

Literatúra

[BSZ] R. Bělohlávek, V. Sklenář, J. Zacpal: Crisply generated fuzzy concepts, in: B. Ganter and R. Godin (Eds.): ICFCA 2005, Lecture Notes in Computer Science 3403, pp. 268–283, Springer-Verlag, Berlin/Heidelberg, 2005.

[BV] R. Bělohlávek, V. Vychodil: Reducing the size of fuzzy concept lattices by hedges. in: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22-25, 2005, Reno (Nevada, USA), pp. 663–668 (proceedings on CD), abstract in printed proceedings, p. 44, ISBN 0-7803-9158-6.

[BJ] S. Ben Yahia, A. Jaoua: Discovering knowledge from fuzzy concept lattice. In: Kandel A., Last M., Bunke H.: Data Mining and Computational Intelligence, 169–190, Physica-Verlag, 2001

[GW] B. Ganter, R. Wille: Formal Concept Analysis, Mathematical Foundation, Springer Verlag 1999, ISBN 3-540-62771-5

[K] S. Krajči: Cluster based efficient generation of fuzzy concepts, Neural Network World 13,5 (2003) 521–530

[RS] M. D. Rice, M. Siff: Clusters, Concepts, and Pseudometrics, Electr. Notes Theor. Comput. Sci. 40: (2000)

Annotation:

Social networks and fuzzy formal concept analysis

The respecting of people relationships is one of the most important factorsx of effective team-building. Therefore a personal manager of social network should know its structure from this point of view too. In our paper we deal with a special social network, namely one secondary grammar school class. We have made the following experiment: Each student characterized his/her relationship to each his/her schoolmate by a value from a given range and we obtained the result table. Then we applied one-sided fuzzy concept lattice including modified Rice & Siff's algorithm – a special data-mining method of formal concept analysis – for this table and obtained the result clusters, i. e. groups of pupils which are sensed by their schoolmates in the same way. In the end of our experiment we interpreted these clusters as near or wider friendships between them. After describing our experiment we discuss possibilities of using of this approach to other social networks, especially to firms. The knowledge obtained in this way can help to a personal manager to compose teams without personal conflicts of their members.