

Metody analýzy dat II

Analýza sítě

OBSAH

1	ÚVOD	4
2	ANALÝZA ZÁKLADNÍCH STRUKTURÁLNÍCH VLASTNOSTÍ SÍTĚ	5
2.1	Náhled na kumulativní grafy všech sítí	6
2.2	Počet vrcholů, hran	7
2.3	Průměrný stupeň, maximální stupeň	7
2.4	Shlukovací koeficient	7
2.5	Souvislost.....	8
2.6	Grafy pro časové snímky	8
2.6.1	1	8
2.6.2	2.....	8
2.6.3	3.....	9
2.6.4	4.....	9
2.6.5	5.....	10
2.6.6	6.....	10
2.6.7	7.....	10
2.7	Vizualizace v Gephi.....	11
2.7.1	1	11
2.7.2	2.....	12
2.7.3	3.....	13
2.7.4	4.....	14
2.7.5	5.....	15
2.7.6	6.....	16
2.7.7	7.....	17
2.8	Analýza vícevrstvých vlastností	18

1 ÚVOD

Datová sada popisuje interakci dvou uživatelů na sociální síti, z online komunity studentů University of California, Irvine. Datová sada zahrnuje uživatele, kteří poslali nebo obdrželi alespoň jednu zprávu.

2 ANALÝZA ZÁKLADNÍCH STRUKTURÁLNÍCH VLASTNOSTÍ SÍTĚ

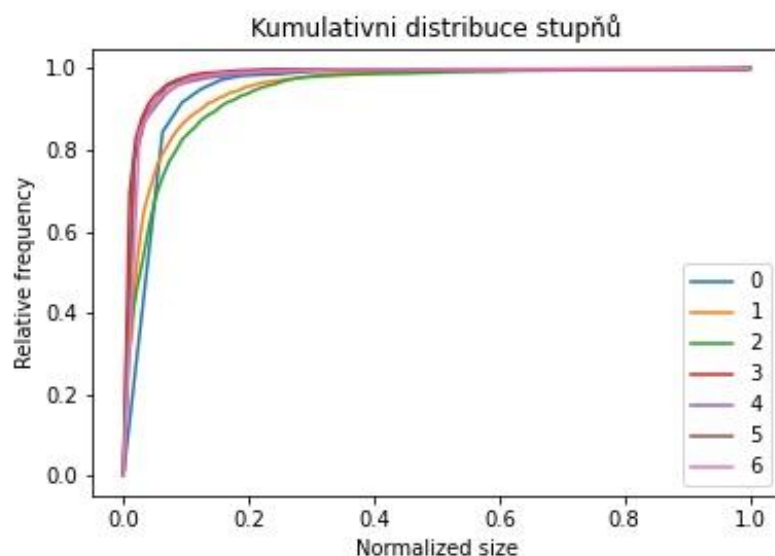
Po detailnější analýze datové sady, bylo zjištěno, že zahrnuté zprávy byly z časového intervalu:

- **Začátek** – středa, 24.3.2004 (2:11:55)
- **Konec** – úterý, 26.10.2010 (07:52:2)

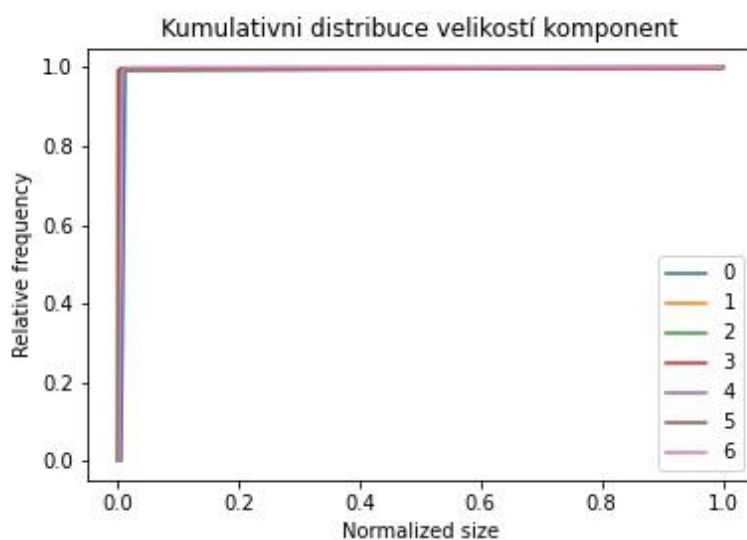
Bylo provedeno rozdělení datové sady, do 7 časových snímků, o stejně dlouhém časovém dění, tudíž (časové snímky byly rozděleny se záměrem trvání jeden měsíc):

1. **Od Wed Mar 24 02:11:55 2004 - Do Sat Apr 24 00:30:38 2004**
2. **Od Sat Apr 24 00:57:38 2004 - Do Mon May 24 21:44:12 2004**
3. **Od Mon May 24 21:48:09 2004 - Do Thu Jun 24 18:43:14 2004**
4. **Od Thu Jun 24 21:25:43 2004 - Do Sun Jul 25 15:52:25 2004**
5. **Od Sun Jul 25 16:22:15 2004 - Do Wed Aug 25 13:23:15 2004**
6. **Od Wed Aug 25 15:03:21 2004 - Do Sat Sep 25 10:18:52 2004**
7. **Od Sat Sep 25 16:29:27 2004 - Do Tue Oct 26 07:52:22 2004**

2.1 Náhled na kumulativní grafy všech sítí



Z následujícího grafu, lze pozorovat, že druhá polovina grafů, reprezentující časová okna, obsahují velké množství izolovaných vrcholů. A tak bychom mohli přepokládat, že v této době v sociální síti nebylo aktivních tolik aktérů. Oproti tomu první tři časová okna by mohla zachycovat aktivnější sociální síť.



Z následujícího grafu, lze pozorovat, že ve všech časových snímcích můžeme očekávat velké množství nezapojených účastníků.

2.2 Počet vrcholů, hran, hustota

	1	2	3	4	5	6	7
Počet vrcholů	1899	<u>1899</u>	1899	1899	1899	1899	1899
Počet hran	558	<u>9317</u>	4764	1086	745	633	392
Hustota	0.00031	0.00517	0.002644	0.000603	0.000413	0.000351	0.000218

V rámci 2. a 3. časového okna lze pozorovat větší hustotu sítě. A tudíž můžeme odhadovat, že se v té době do komunikace zapojilo větší množství účastníků, než je v ostatních měsících zvykem.

2.3 Průměrný stupeň, maximální stupeň

	1	2	3	4	5	6	7
Průměrný stupeň	0.59	<u>9.81</u>	5.02	1.14	0.78	0.67	0.41
Maximální stupeň	31	<u>216</u>	95	100	57	57	38

V druhém měsíci můžeme pozorovat, že se vyskytuje student se stupněm 216. Značí, že většina studentů se z nějakého důvodu více shlukovala kolem určitých jedinců.

2.4 Shlukovací koeficient

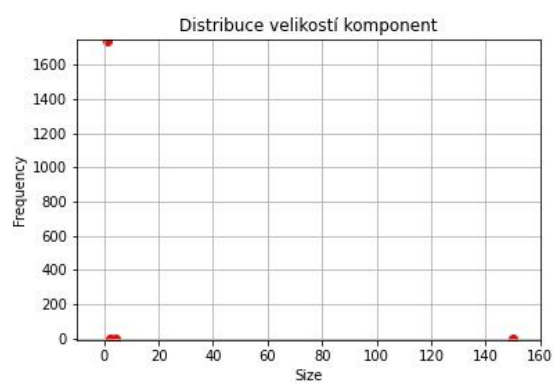
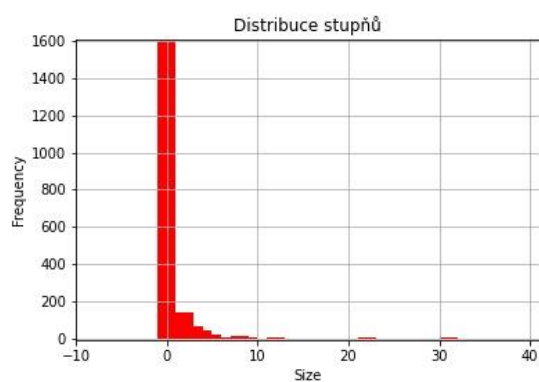
	1	2	3	4	5	6	7
Průměrný shlukovací koeficient	0.001683	<u>0.078686</u>	0.032086	0.015538	0.007304	0.004322	0.000950031

2.5 Souvislost

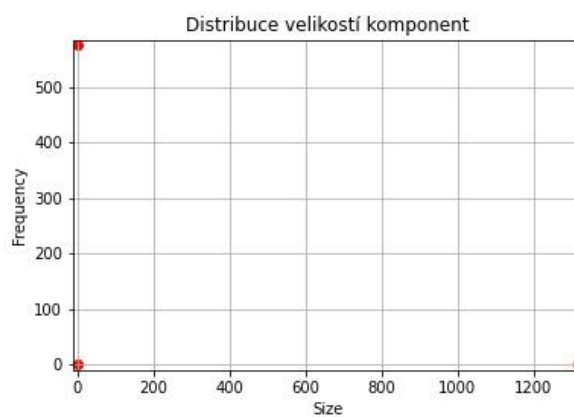
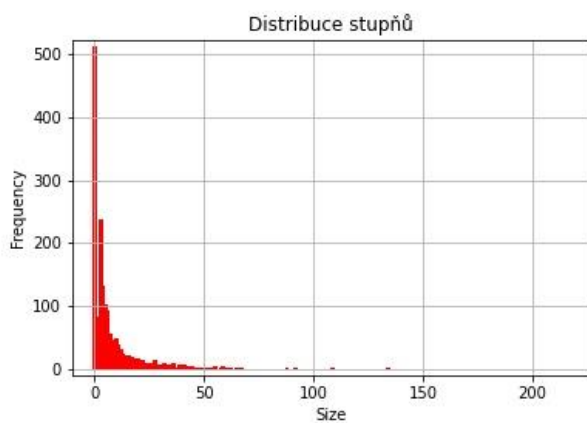
	1	2	3	4	5	6	7
Počet komponent souvislosti	1745	<u>579</u>	708	1338	1495	1490	1616

2.6 Grafy pro časové snímky

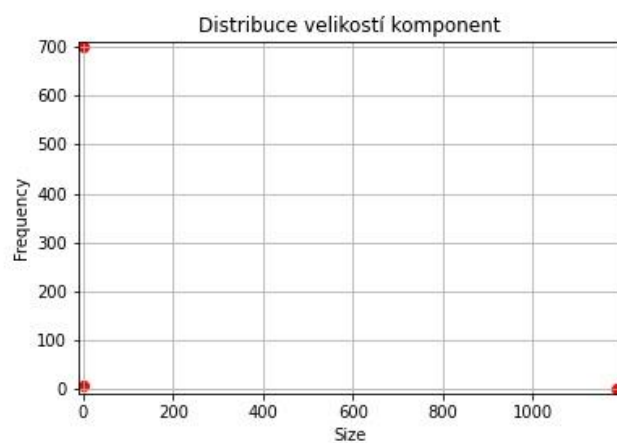
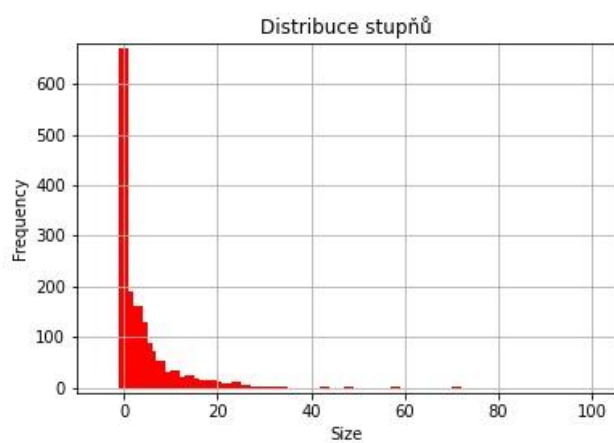
2.6.1 1



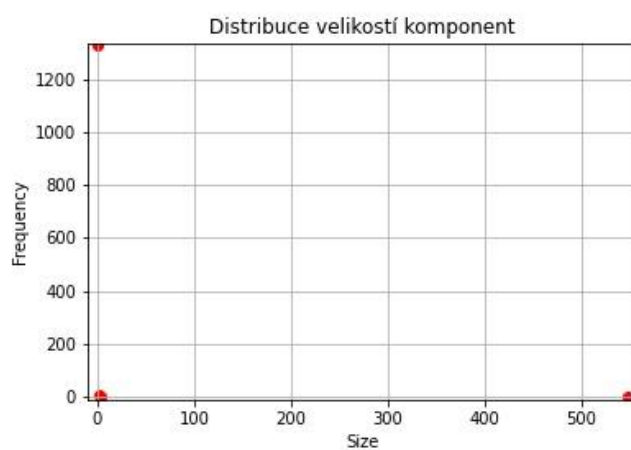
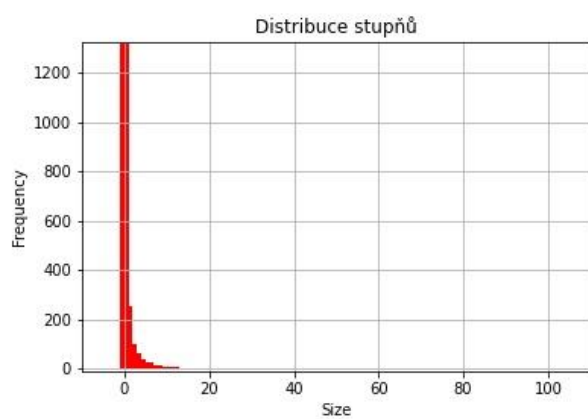
2.6.2 2



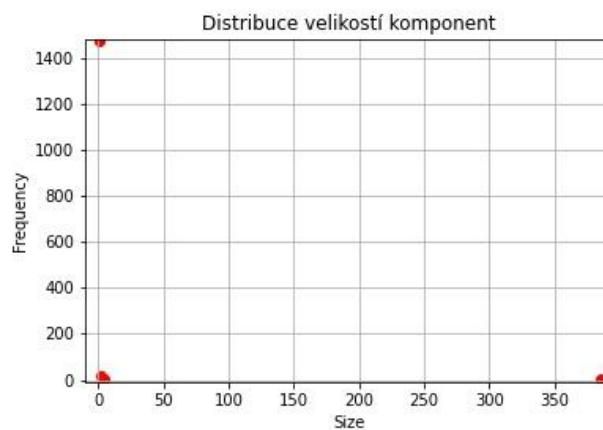
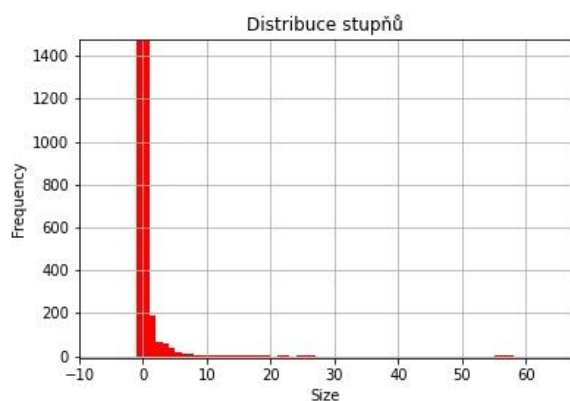
2.6.3 3



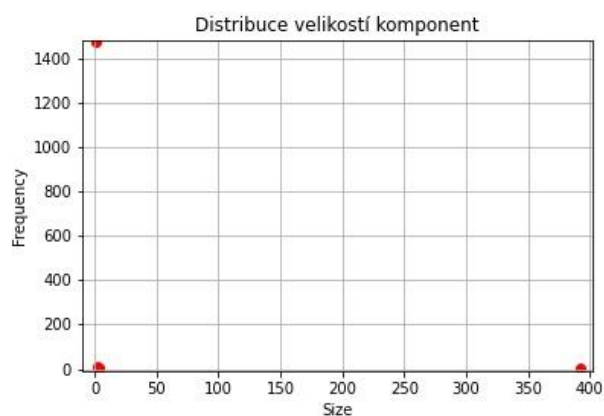
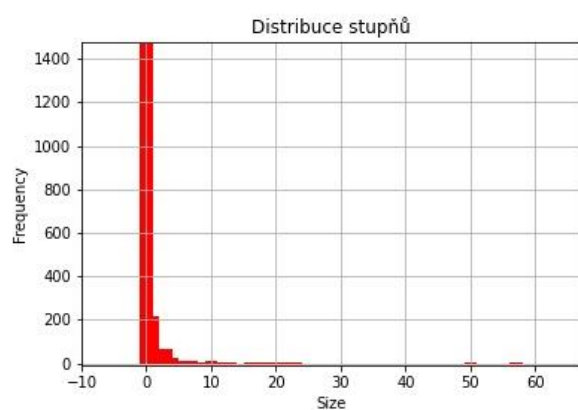
2.6.4 4



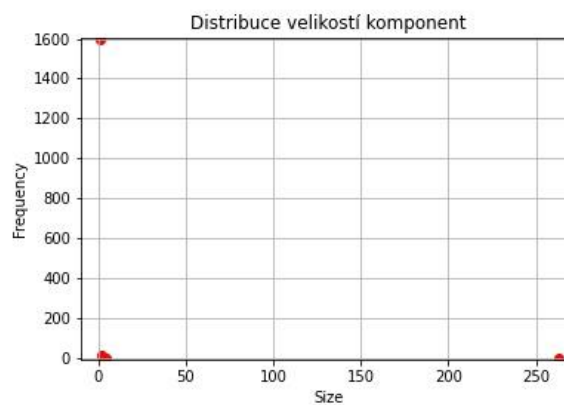
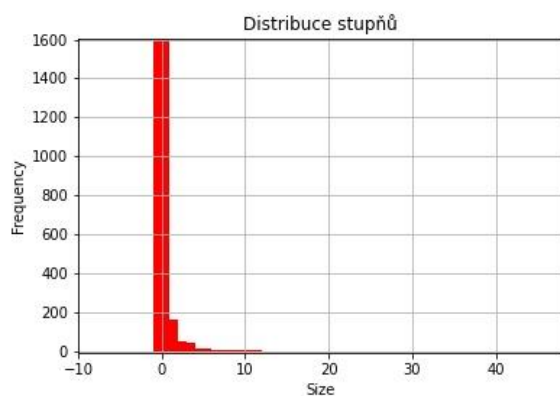
2.6.5 5



2.6.6 6



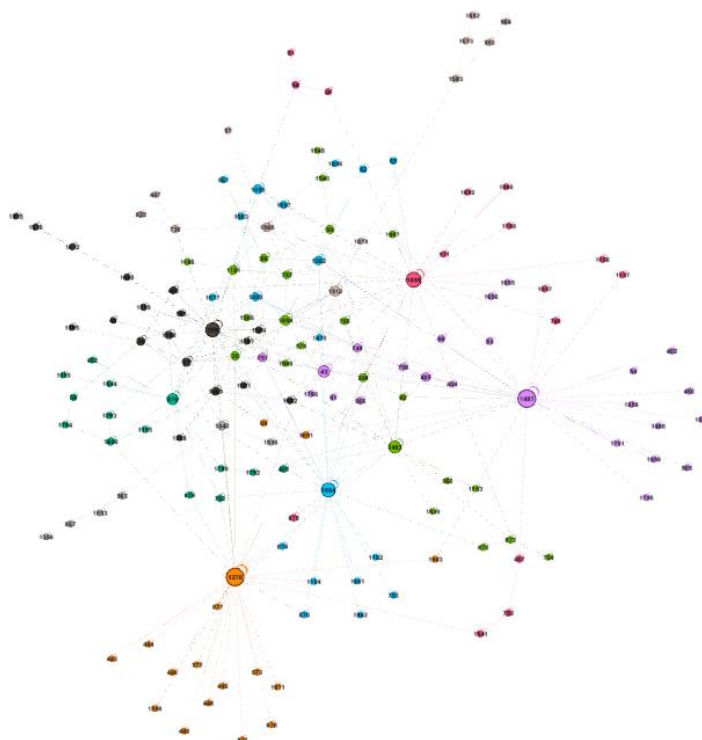
2.6.7 7



2.7 Vizualizace v Gephi

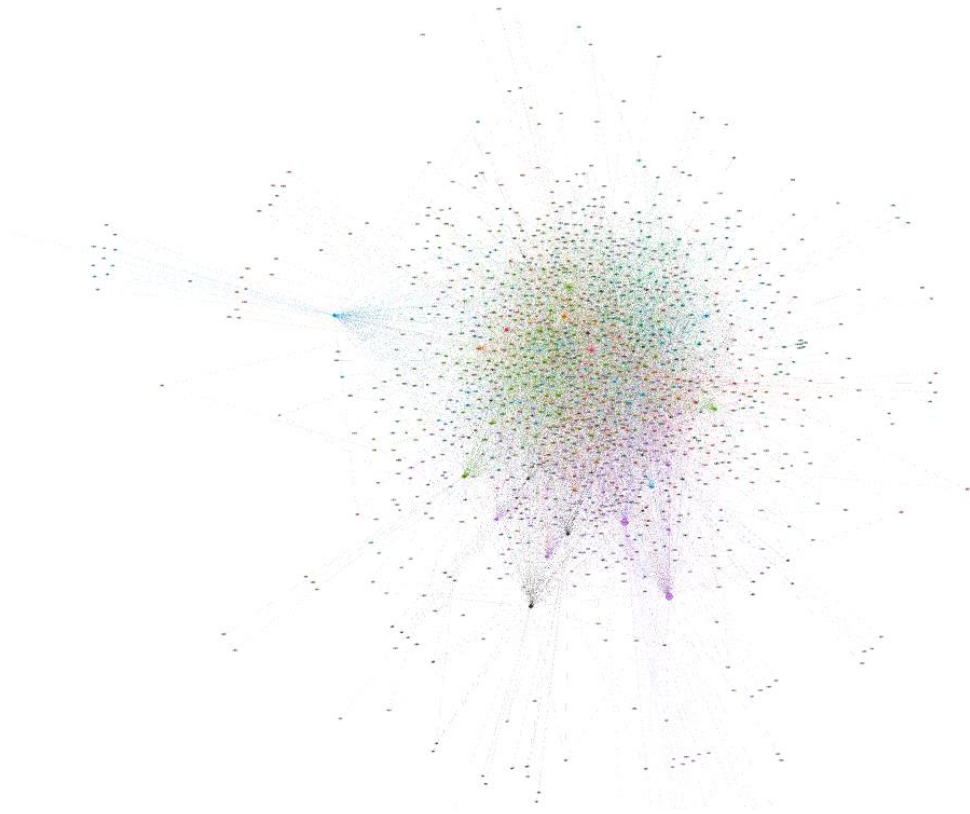
2.7.1 1

1007

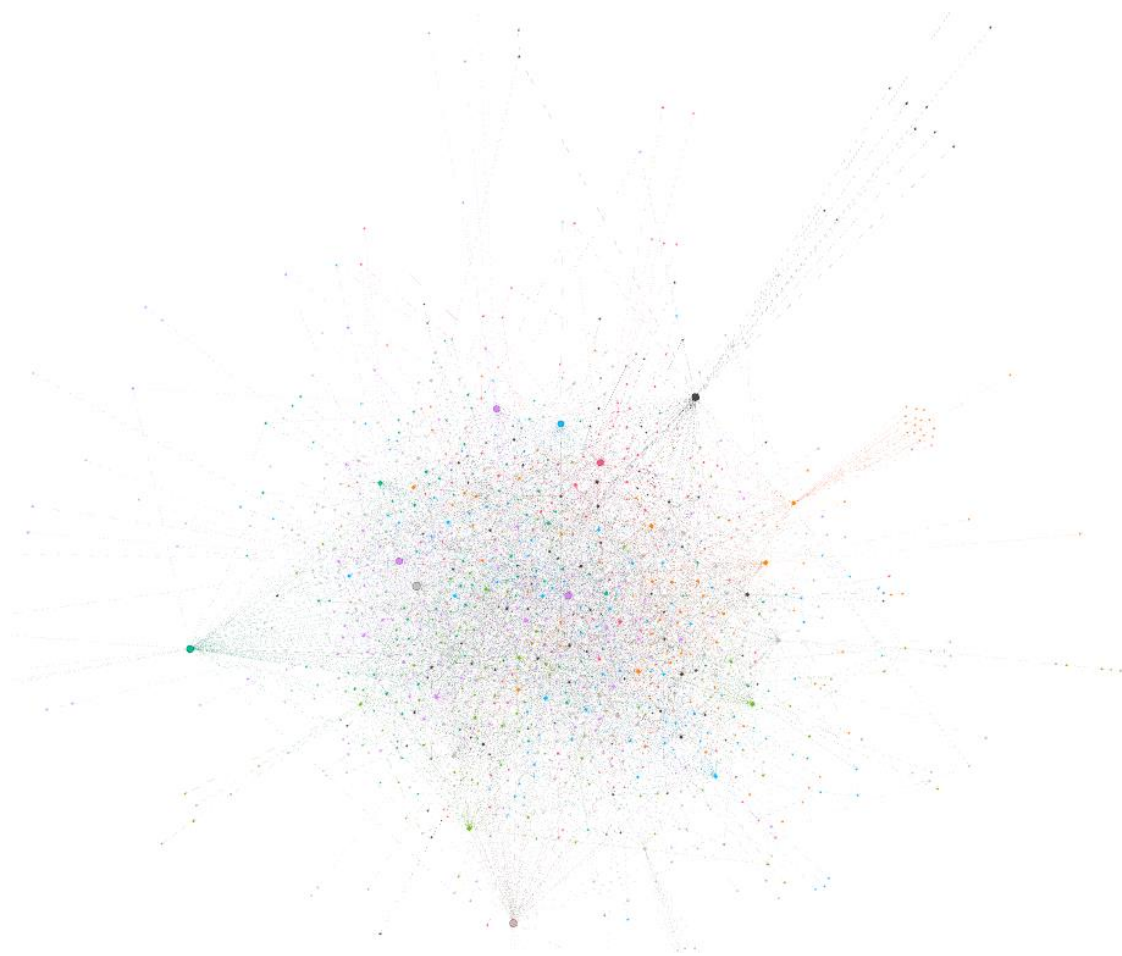


1004

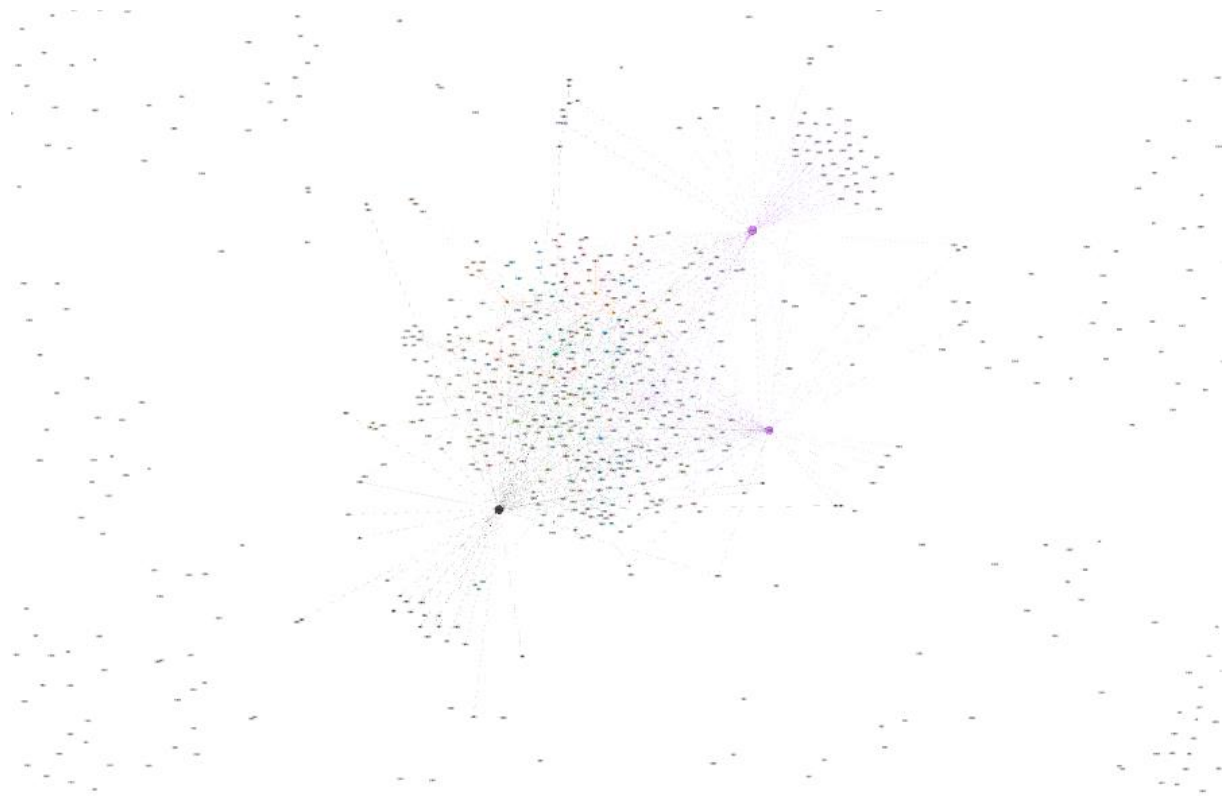
2.7.2 2



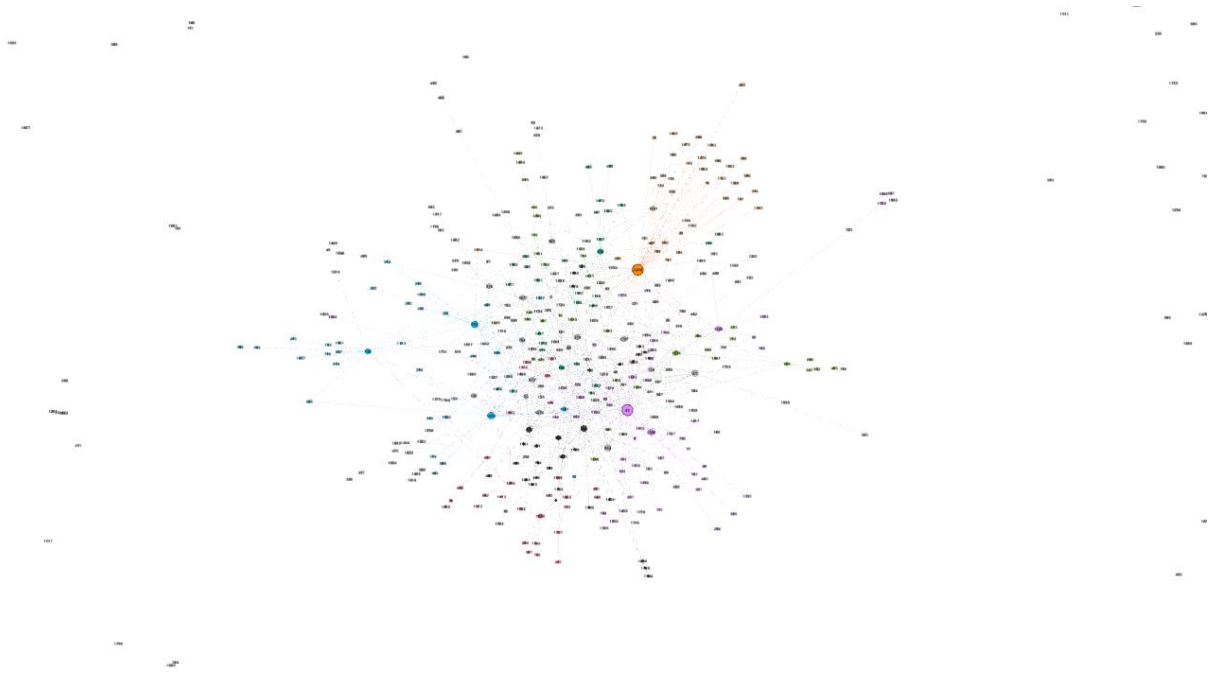
2.7.3 3



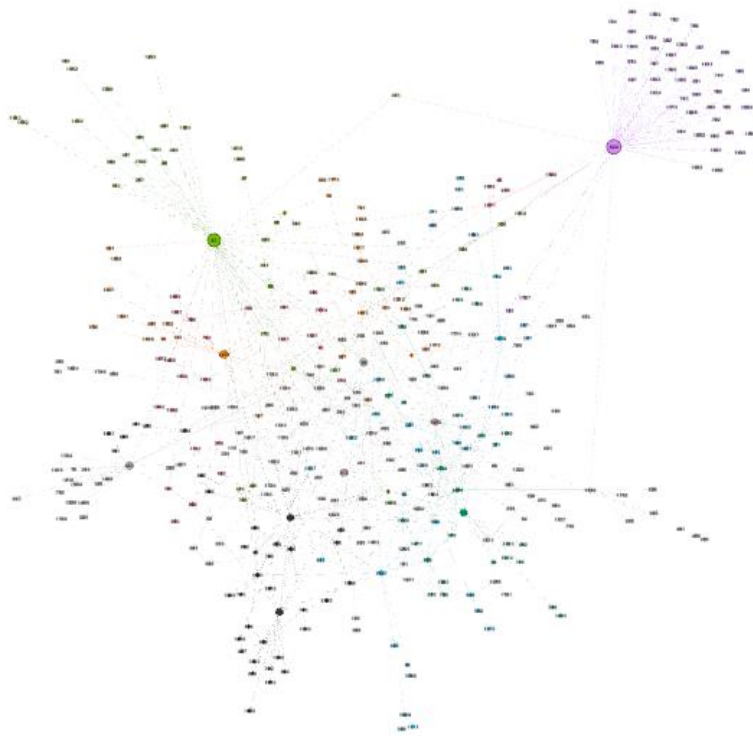
2.7.4 4



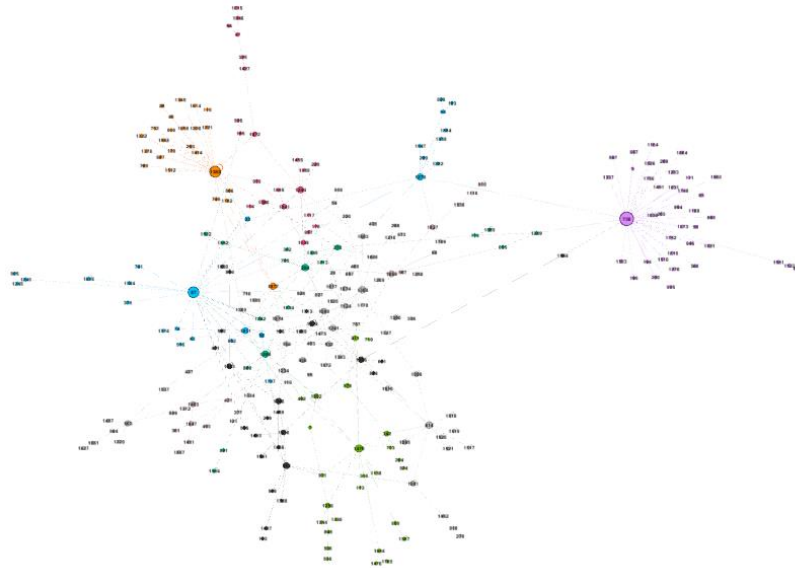
2.7.5 5



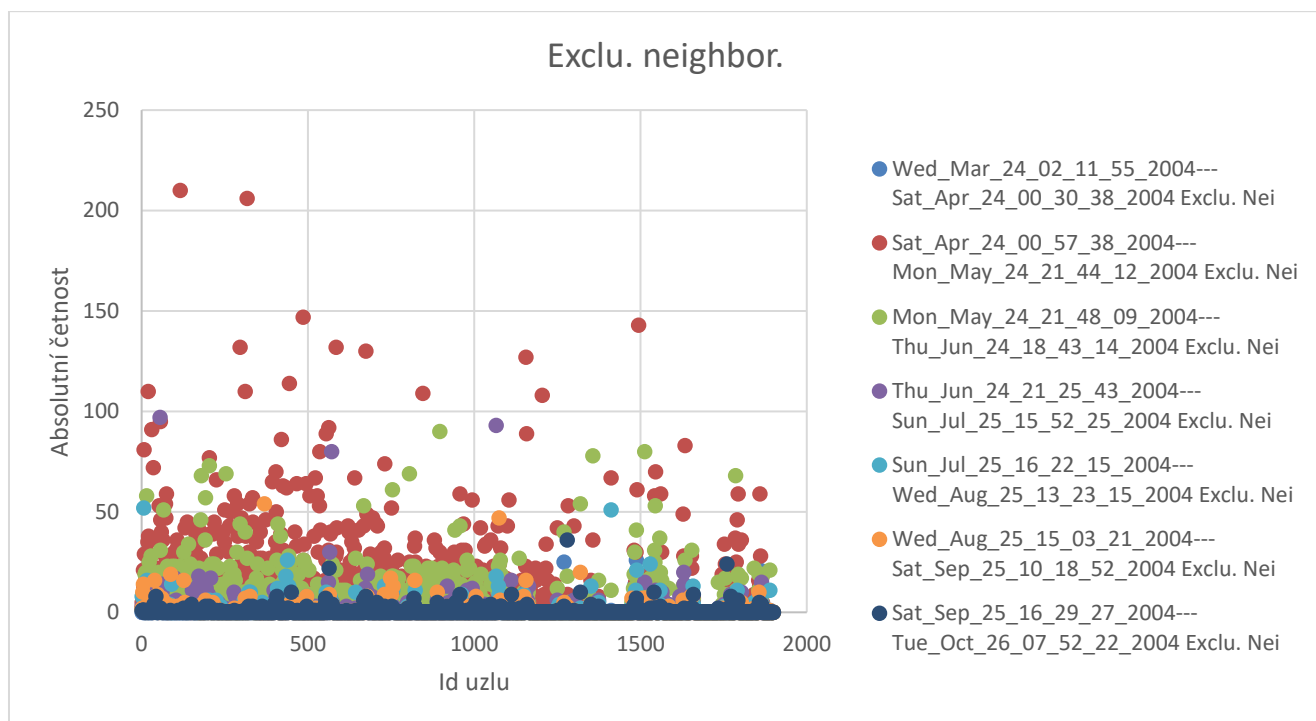
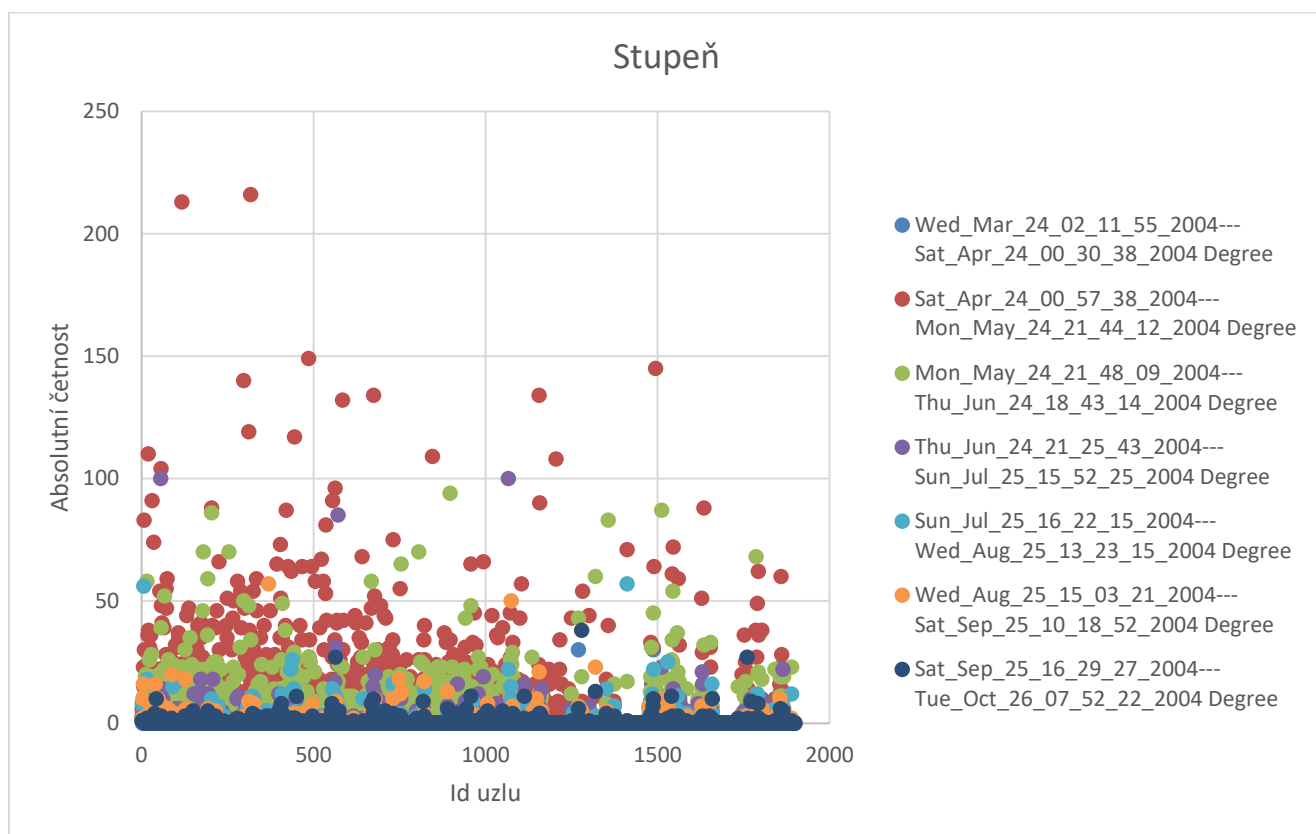
2.7.6 6

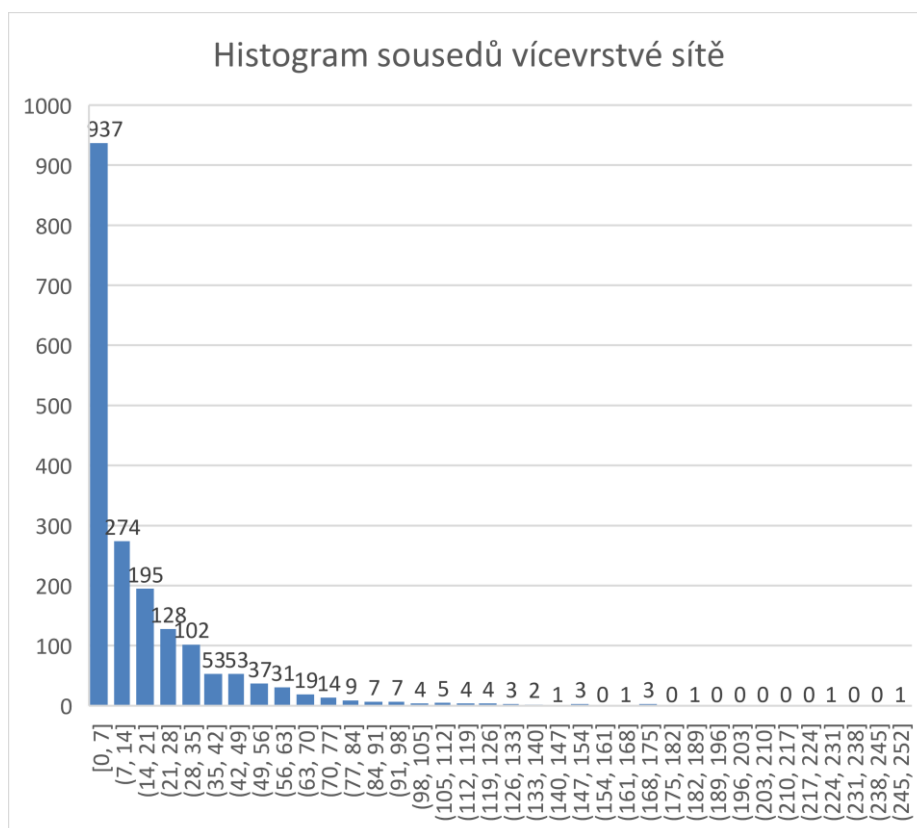
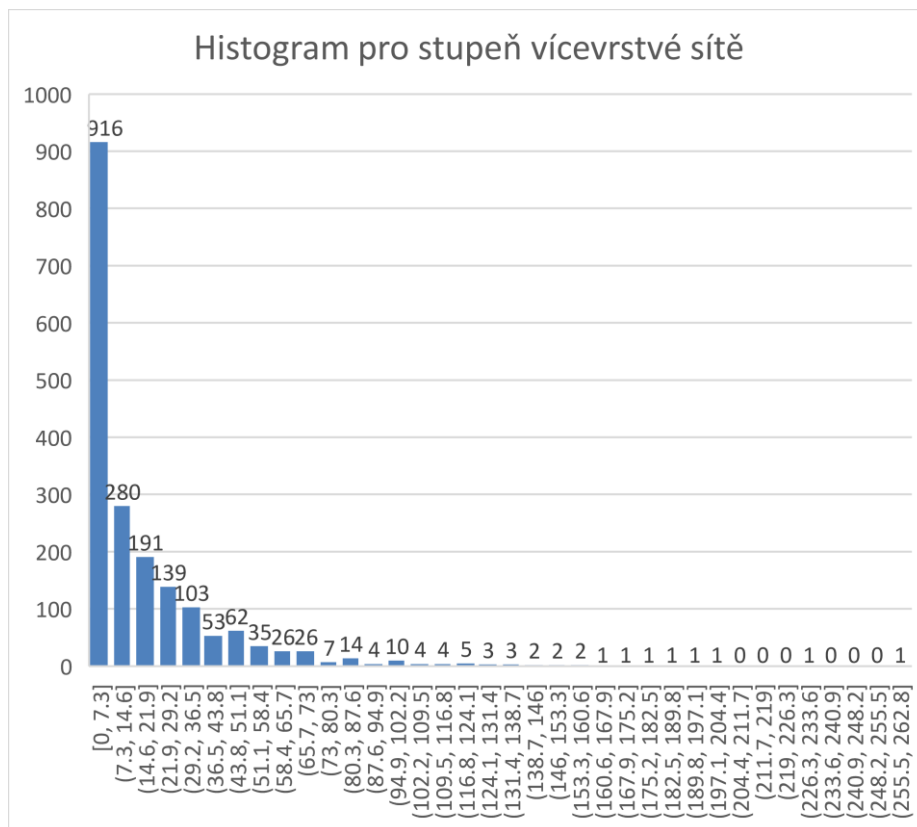


2.7.7 7



2.8 Analýza vícevrstvých vlastností





3 SHRnutí

Z předchozích tabulek a grafů šlo pozorovat vyšší aktivitu v rámci sociální sítě v 3 časových snímcích. Tudíž šlo o měsíce **duben, květen, červen, červenec**. Především pak nejvyšší aktivitu šlo pozorvat v druhém časovém snímku, znázorňující aktivitu od 24. dubna do 24. května. Současně v této době, docházelo k zapojení takových účastníků, kteří v porovnání s ostatními časovými snímky nekomunikovali.

V rámci nejaktivnějšího časového okénka lze pozorovat velice souvislý graf.

V každém časovém okně, lze pozorovat uzly, které komunikují s větším počtem ostatních studentů. Tudíž lze na tento stav nahlížet, že se studenti shlukují kolem určitého jednoho studenta. Struktura je lehce podobající se ego síti.

Ve všech časových oknech, šlo vždy pozorovat spoustu izolovaných vrcholů, které znázorňují účastníky, kteří ačkoliv byli v sociální síti, tak se nezapojovali (zaslání nebo obdržení zprávy).