

Modelování dynamiky informačních toků v sociálních sítích

Jiří Jelínek¹

¹ Jihočeská univerzita v Českých Budějovicích, Přírodovědecká fakulta, Branišovská 31,
370 05 České Budějovice, Česká Republika
jelinek.fm@gmail.com

Abstrakt. Sociální sítě patří v současné době díky své popularitě mezi uživateli k nejvíce studovaným strukturám v oblasti výměny informací a znalostí. Tyto struktury jsou velmi dobře popsány z hlediska jejich statického chování, tento příspěvek se snaží popsat rovněž dynamiku šíření informací a znalostí v těchto sítích. Proto bude představen heuristický model chování jedince v síti založený na jeho potřebě získávání informací a znalostí pro řešení konkrétních situací. Z něho vychází agent-based model celé sociální sítě, který byl prakticky implementován a prezentovány budou rovněž jeho výstupy při simulaci vybraných úloh testujících dynamiku sociální sítě.

Klíčová slova: sociální sítě, AB simulace, šíření informací

1 Úvod

Sociální sítě patří v současné době k nejvíce studovaným strukturám v oblasti výměny informací a znalostí. Tyto struktury jsou velmi dobře popsány z hlediska jejich statického chování, náš příspěvek se snaží popsat rovněž dynamiku šíření informací v jejich rámci.

Sociální sítě je možné si představit jako grafické struktury, jejichž uzly tvoří objekty či individuality (lidé) a jejichž hrany představují nějaký typ propojení mezi těmito objekty. Charakter tohoto propojení může být různý, vždy však vyjadřuje určitou vazbu či komunikaci mezi propojovanými uzly.

Konstrukce a identifikace sociálních sítí vychází především z informací získaných od jednotlivců a také např. ze záznamů vzájemné komunikace. Typickými systémy pro podporu sociálních sítí jsou on-line služby jako např. Facebook, MySpace, atd.

Tento příspěvek si klade za cíl ukázat jednu z možných cest modelování a simulace dynamiky šíření informací a znalostí v sociální síti a na experimentálních výsledcích ukázat některé jevy, které se v dynamice sítí projevují.

Model může být využit pro zkoumání dynamiky v sítích různého rozsahu i zaměření vytvořených za účelem přenosu informací (např. podnikové sítě, on-line služby, lokální struktury zaměřené na řešení běžných životních situací, atd.), neomezuje se pouze na on-line služby a elektronický přenos informací.

2 Současný stav

Řešení problémů s pomocí simulačních technik se zejména v poslední době stalo velmi populárním postupem. Již od 60. let minulého století se používá přístup založený na tzv. systémové dynamice (SD). Ten byl nejprve označován jako „Industrial Dynamics“ a byl poprvé postulován Jay Forresterem v roce 1961 [5]. Druhým přístupem je modelování dějů na základě výskytu konkrétních situací (DES – Discrete Event Simulation). V tomto případě je modelovaný systém popsán pomocí stavů a událostí, které tyto stavy ovlivňují a mění. Třetím přístupem je založení simulace na modelování menších často se vyskytujících objektů, které svým chováním a vzájemnými interakcemi vytvářejí globální charakteristiku úlohy. Uvedený přístup se označuje jako agentový (AB – agent based). Je jasné, že tento přístup je možný pouze tam, kde lze v úloze identifikovat mikroprvky vyskytující se ve větším počtu jedinců, což je právě případ sociálních sítí. Pro modelování na úrovni agentů je možno využít jak SD, tak DES přístup, případně AB přístup na nižší hierarchické úrovni.

Porovnání SD, DES a AB přístupu s hlavním důrazem na poslední z uvedených lze nalézt např. v [3]. Výběr konkrétního přístupu k návrhu modelu je dán rovněž dostupností vstupních dat (globálních či lokálních). U AB přístupu jsou preferována lokální data o chování jedinců, případně globální pravděpodobnostní data umožňující parametry jedince odvodit. Oblibu AB přístupu zvyšuje i množství a kvalita nástrojů pro AB modelování, další informace lze nalézt např. v [9].

Problematikou sociálních sítí se zabývá celá řada publikací, přičemž v poslední době se objevují rovněž příspěvky zaměřené právě do oblasti dynamiky těchto sítí a šíření dat informací v těchto sítích. Přehled o technikách analýzy sociálních sítí je možné nalézt např. v [12].

Některé ze směrů zkoumání dynamiky sociálních sítí jsou shrnuty v [2]. Zmíněn je statisticko-mechanický model považující síť za fyzikální systém, u kterého hledáme zákonitosti chování. Druhým je tzv. algoritmický přístup vycházející z informačních sítí. V [2] je rovněž prezentován nový přístup založený na monitorování dynamiky sítě pro seskupování uživatelů podle jejich vzájemné podobnosti, přičemž vytvořené skupiny jsou pak dále analyzovány. Stále častěji se objevují i modely techničtější zaměřené a vycházející ze simulačních technik, zejména multiagentního přístupu.

Agentovému přístupu a jeho užití v analýze sociálních sítí (e-mailových) se věnuje i [6]. Zde je možno seznámit se i s různými modely růstu sociálních sítí, prezentovaný model pak vychází z modelu společných přátel a preferenčních podmínek pro volbu partnera. Model agenta je zde založen na dvou základních parametrech – přitažlivosti a zájmu. Podobně jako ve zde prezentovaném modelu jsou tyto parametry voleny na pravděpodobnostním principu.

Myšlenka specifikace parametru důvěry, a to i mezi nepropojenými agenty v síti je prezentována v [11], kde je představena nová metrika pro výpočet důvěry vycházející z technik jako je např. PageRank. Popsaný model chápe hrany v síti jako vyjádření důvěry (určitého ratingu) mezi agenty, agent pak komunikuje pouze s důvěryhodným protějškem. Prezentovaný příspěvek byl tímto přístupem inspirován.

V [4] se autoři zabývají analýzou dynamiky šíření informací v populární službě Flickr a ukazují, jak jsou informace sociální sítí této služby přenášeny. Podobná studie

v širším měřítku zahrnující služby Flickr, YouTube, LiveJournal a Orkut je rovněž k dispozici [7].

Možnostmi predikce chování individuí v sociální síti při transferu informací se věnuje [10]. Snahou je predikovat jak cestu informace, tak její obsahové zaměření.

Šířením myšlenek v online sítích se zabývá i [1], kde je použito rovněž několika měr, např. akceptovatelnost, přenositelnost či dosažitelnost. I tímto přístupem byl ovlivněn navrhovaný model.

Nalézt lze i zdroje zaměřené na simulaci rozhodovacích procesů jedinců (např. [8]), které se zaměřením příspěvku také částečně souvisí.

Žádný z uvedených zdrojů však nepracuje s modelem, který je popsán v následujících odstavcích, vycházejícím z potřeby získat informace k nalezení řešení daného problému.

3 Modelování sociální sítě a jejích prvků

Pro modelování dynamiky sociální sítě byl zvolen AB přístup vycházející z koexistence a vzájemné komunikace elementárních objektů (agentů). Na úrovni agenta je pak pro popis chování užitá technika DES.

Pro účely simulace dynamiky sítě předpokládáme existenci agentů v daném prostoru, kteří jsou vystavováni „životním“ situacím, vyžadujícím jejich rozhodnutí (řešení situace). Agenti se snaží o co nejlepší rozhodnutí. Pro jeho přijetí vychází z vlastních informací a schopností, ale snaží se zkoumat i řešení téže situace přijatá v minulosti ostatními agenty. Model vychází z předpokladu uzavřeného světa uplatněného na počet uzlů v síti a množinu možných situací, který je v něm konstantní a v čase neměnný. Tento předpoklad neodpovídá reálnému stavu, je však přijatelným zjednodušením. Naším cílem je především zkoumat, jakým způsobem se budou vyvíjet znalosti o řešení příslušných situací u jednotlivých agentů i v síti jako celku (dynamika vazeb).

Modelování samotných situací není stěžejní otázkou, pro účely modelu jde pouze o vygenerování čísla dané situace, přičemž „řešením“ každé možné situace j z celkového počtu N agentem i je číslo s_{ij} z intervalu $\langle 0;1 \rangle$. Podstatné pro další postup je pouze to, že číslo s_{ij} lze použít jako míru umožňující stanovit kvalitu řešení ($1 =$ maximální kvalita, tedy optimální řešení).

3.1 Podrobný popis modelu

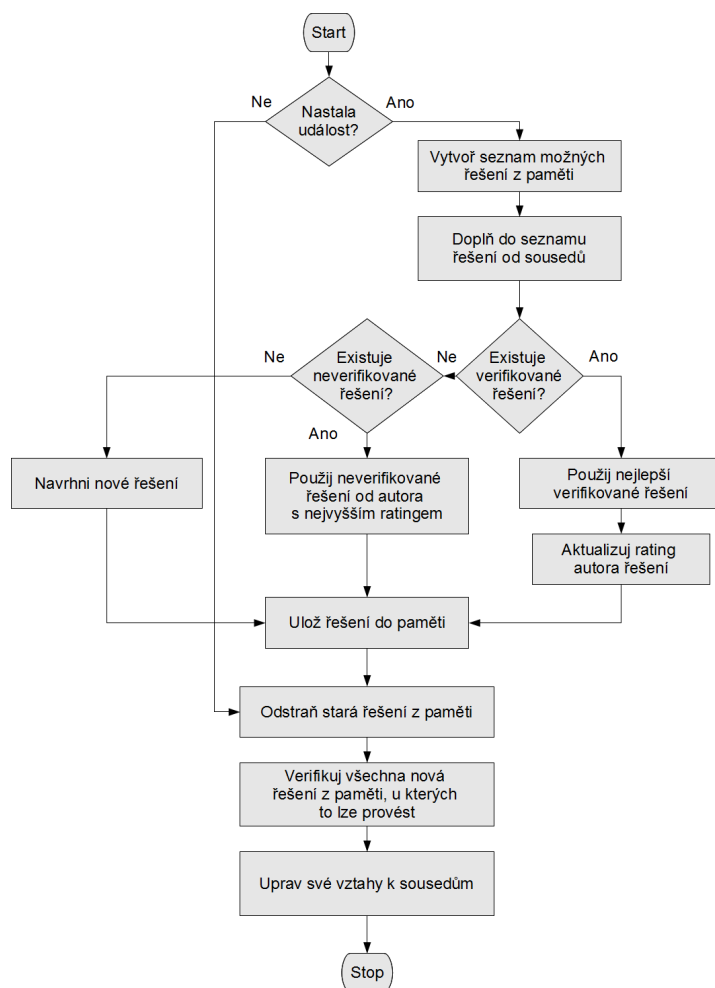
Každý z agentů je vybaven vlastní pamětí zachycující předchozí přijatá rozhodnutí v různých situacích a tvořící základ jednoduchého CBR (case based reasoning) systému. V případě, že v paměti se řešení k uvedené situaci již vyskytují, tvoří tato řešení bázi k přijetí aktuálního řešení.

Pokud je dané situaci agent vystaven poprvé, pak bez existence sociální sítě generuje vlastní řešení na základě postupů modelujících úroveň vzdělání jedince. To je simulováno schopností agenta generovat řešení situace s_{ij} s kvalitou v definovaném intervalu hodnot. Použito je zde rovnoměrné rozložení pravděpodobnosti. Pokud je agent součástí sítě, snaží se nalézt pro danou situaci řešení nejen ve vlastní paměti, ale

i dotazem na své sousedy. Při komunikaci je respektována různá ochota agentů komunikovat, a to pomocí míry *commRate* z intervalu $<0;1>$, nastavené individuálně pro každého agenta, a míry *acceptRate* (rovněž z intervalu $<0;1>$) vyjadřující ochotu dotazovaného agenta přijmout dotaz a odpovědět na něj.

V procesu hledání řešení situace hraje zásadní úlohu zpětná vazba, tedy možnost verifikace kvality příslušného řešení. Bez ní není agent schopen určit kvalitu navrhovaného či přijatého řešení a nelze tedy preferovat kvalitnější rozhodnutí. Model respektuje fakt, že řešení nemusí být verifikovatelné ihned po svém přijetí, ale až po definované době.

Přijaté řešení si agent ukládá do paměti spolu s údajem, kým bylo vytvořeno. Tvůrcem řešení přitom nemusí být nutně agent, od kterého je získáno, i ten jej mohl převzít od jiného jedince.



Obr. 1. Diagram činnosti agenta v každém časovém kroku.

Informace o autorech řešení jsou využity pro rating agentů v síti, a to nejen bezprostředních sousedů agenta, ale v rámci celé sociální sítě postupným přenosem informací po vazbách mezi jednotlivými agenty. Autor řešení se přidává do seznamu hodnocených agentů (partnerů) a jeho rating se agentovi upravuje v okamžiku verifikace jím navrhovaného řešení podle vzorce (1). V něm je k_r koeficient změny ratingu (v experimentech nastaven na $k_r = 0,1$), s_{ij} kvalita řešení situace j generovaného agentem i a $r_{i,new}$ a $r_{i,old}$ nová a původní hodnota ratingu agenta.

$$r_{i,new} = r_{i,old} + k_r (s_{ij} - r_{i,old}), \quad (1)$$

Rating je využit v situaci, kdy není k dispozici verifikované řešení dané situace získané ze sítě či z paměti agenta. Jedná se o určitý preferenční model používaný i při vytváření a mazání vazeb v síti. Popsaný mechanismus je jednou z možností, jak do modelu implementovat princip lokální důvěry popisující orientovanou vazbu mezi dvěma agenty. S postupem simulace agent získává informace o dalších agentech v rámci sítě, což má za následek modifikaci vazeb. Délku seznamu partnerů je možné definovat v rámci zadání, vymazávání jsou pak agenti s nejnižším ratingem.

Celkový model chování agenta je uveden na obr. 1. Situací uvedených v zadání existuje uzavřená množina, její elementy se vyskytují s pravděpodobností, kterou lze nastavit. V paměti agenta je implementován i proces zapomínání, po určité době (parametr *forgetTime* každého agenta) jsou z paměti vymazávány údaje o starých řešeních.

3.2 Způsob realizace

Podle výše popsaného simulačního modelu sociální sítě byla vytvořena jeho implementace.

Navržené prostředí umožňuje před započítím simulace nastavit základní globální parametry, podle kterých jsou potom vygenerovány konkrétní hodnoty charakterizující každého agenta.

Abychom byli schopni charakterizovat objem znalostí, které daný agent získal (jeho schopnost kvalitně reagovat na předkládané situace či události) je pro každého agenta i vypočítán koeficient jeho kvality q_i , který je získán podle (2), kde N je celkový počet situací a s_{ij} kvalita každého řešení situace j tímto agentem. Pokud řešení pro situaci j není k dispozici, bere se hodnota s_{ij} jako nulová.

$$q_i = \frac{1}{N} \sum_{j=1}^N \max(s_{ij}) \quad (2)$$

Definována byla i celková kvalita sítě z hlediska distribuovanosti znalostí Q , a to podle (3), kde M je celkový počet agentů v síti.

$$Q = \frac{1}{M} \sum_{i=1}^M q_i \quad (3)$$

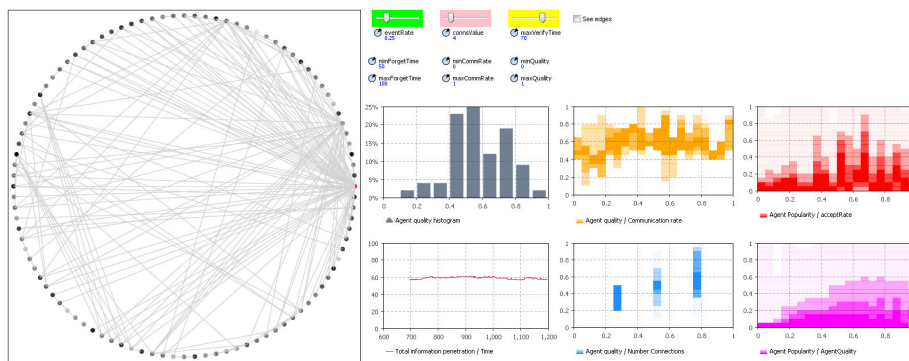
Po zahájení simulace je síť nastavena tak, že agenti mají vazby pouze na své dva sousedy vybrané na základě identifikačního čísla (id agenta zvětšené / zmenšené o 1, u prvního a posledního agenta v číselné řadě jsou sousedy i koncoví agenti z druhého konce řady – cyklické uspořádání).

$$p_i = \frac{1}{M} \sum_{j=1}^M r_{ji} \quad (4)$$

Pro posuzování množství odkazů a ratingu konkrétního agenta i byla definována i míra jeho „oblíby“ a to podle (4), kde p_i je míra popularity agenta a r_{ji} rating agenta i u agenta j (pokud není zvolený agent i u některého jedince j hodnocen, je nastaveno $r_{ji}=0$). Na tomto místě je potřeba upozornit, že hodnocení provádí vybraný agent i sám na sobě, tedy že smysl má i rating r_{ii} .

4 Výsledky experimentů

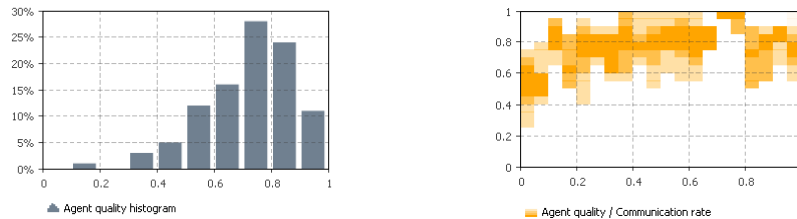
Vytvořený simulační model umožňuje shromažďovat během simulace význačná data o chování agentů. Základním údajem je pak koeficient q_i , rating agentů a počet vazeb agenta. Sledovat lze rovněž celkovou kvalitu sítě Q a další údaje. Ukázky z prováděných experimentů jsou uvedeny na následujících obrázcích. Všechny byly uskutečněny na síti o 100 agentech s 10 situacemi vyžadujícími řešení.



Obr. 2. Prostředí aplikace pro simulaci dynamiky sociálních sítí.

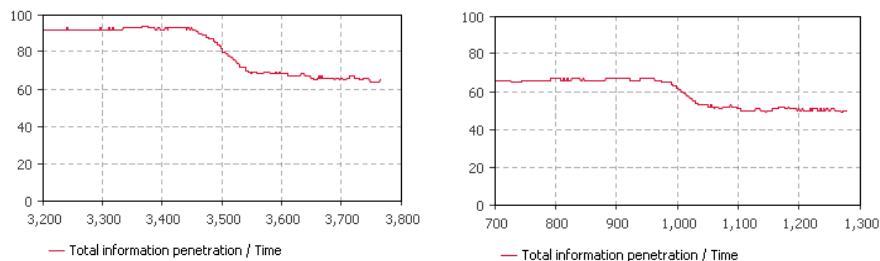
Obrázek 2 představuje navržené simulační prostředí. V jeho levé části je zobrazena samotná sociální síť a vazby v ní. Na začátku lze iniciovat seznam partnerů, v tomto případě jimi byli nejbližší sousedé. S postupem simulace dochází k vytváření dlouhých vazeb zohledňujících kvalitu řešení vytvářených příslušnými agenty. V síti se tak projevuje snaha jejich členů rafinovat své vztahy a vazby tak, aby co nejlépe pokrývaly jejich potřeby. Postupem času se tak jedinec napojuje na stále kvalitnější partnery, v síti vznikají určitá centra (agenti generující nejlepší řešení).

V pravé části se nacházejí ukazatele hlavních parametrů modelu, přičemž hodnoty některých lze měnit v průběhu simulace. V pravé části jsou rovněž zobrazovány grafy zachycující vývoj nebo aktuální stav v síti, z nichž některé jsou popsány dále.



Obr. 3, 4. Grafy zachycující hodnoty kvality agentů q_i v síti

Na obrázku 3 je histogram aktuálního rozložení kvality agentů. Na ose y je uvedeno procentní zastoupení agentů s danou hodnotou q_i . Na obr. 4 je histogram závislosti kvality q_i agenta na jeho ochotě komunikovat. Na ose x je uvedena normovaná míra ochoty ke komunikaci, na ose y pak kvalita agenta q_i . Počty agentů v jednotlivých oblastech jsou vyjádřeny sytostí zabarvení (sytější barva = více agentů v daném poli). Je zřejmé, že při nízké ochotě ke komunikaci je kvalita rozhodnutí agenta horší.



Obr. 5, 6. Vývoj parametru Q v dané sociální síti

Grafy vývoje celkové kvality sítě Q jsou na obrázcích 5 a 6. Osa x zachycuje čas, osa y procentní hodnotu celkové kvality Q . Maximální hodnota 100% je stav, kdy každý agent v síti dokáže ideálně reagovat na všechny možné situace. Obrázky ukazují výsledek experimentu, kdy se projevily vliv sociální sítě na kvalitu agentů. Obrázek 5 ukazuje chování funkční sociální sítě. Agenti v ní dosahují vyšších hodnot kvality než v síti, ve které jsou vzájemné vazby potlačeny. V obou případech pak byla skokově snížena hodnota pravděpodobnosti výskytu situací z 0,5 na 0,2 (střed obrázku), což vede ke zvýšení vlivu zapomínání (agenti mají méně podnětů k obnově svých znalostí). U propojené sítě na obr. 5 její kvalita klesne, ale nedosáhne hodnot z obrázku 6, ilustrujícího chování jedinců bez vazeb v síti. Míra poklesu je však v obou případech srovnatelná a je patrně dána nastavením faktoru zapomínání.

S navrženým modelem byly a jsou prováděny i další experimenty zaměřené na zkoumání chování sítě. Experimentálně bylo např. zjištěno, že v tomto modelu obliba agenta není v ustáleném stavu závislá na hodnotě jeho *acceptRate*, tedy na jeho ochotě odpovídat na dotazy se sítě.

Velmi zajímavé výsledky byly získány při zkoumání celkové kvality sítě Q v závislosti na délce seznamu partnerů, se kterými daný agent komunikuje a od kterých tedy získává možná řešení situací. V ustáleném stavu bylo možno omezit

počet partnerů z původních 20 až na 1, přičemž celková kvalita sítě Q se prakticky nezměnila. Agenti tedy byli schopni vytipovat pomocí ratingu partnera poskytujícího trvale nejlepší řešení a pro přijetí kvalitního řešení jim stačilo komunikovat pouze s ním. Obdobné chování lze vysledovat i v „reálných“ strukturách – vyhledáváme jedince s velmi dobrými znalostmi, od kterých tyto znalosti čerpáme.

Rovněž se ukázalo, že v takto modelované sociální síti může dojít k rozdělení agentů na několik navzájem nekomunikujících skupin. Tento jev nenastane, pokud agenti projeví určitou „iniciativu“ při obnovování svého seznamu partnerů (např. s využitím modelu vyhledávacích služeb, kde je seznam partnerů rozšířen o nalezené jedince). I tento závěr odpovídá realitě – uzavřená komunita se může dostat do separace, která pak ovlivňuje kvalitu reakce jejích členů z důvodu omezeného přístupu k informacím. V modelu by toto chování mohlo být realizováno doplněním náhodného agenta do seznamu sousedů, což by vyjadřovalo jistý „náhodný kontakt“ mezi agenty v síti.

5 Závěr a další postup

V tomto příspěvku byl představen model pro simulaci dynamiky přenosu informací a znalostí v rámci sociální sítě. Na provedených experimentech bylo ukázáno, že model díky komplexní simulaci chování agenta poskytuje zajímavé výstupy, které jsou srovnatelné s chováním jedinců v reálném světě.

Představený model je neustále rozšiřován a upravován. Další práce na něm se zaměří na zkoumání vlivu počtu situací a velikosti sítě na kvalitu nalezených řešení a také na rozšíření funkčnosti modelu eliminací předpokladu konstantního počtu uzlů a situací. Neustále je rovněž hledán vhodný uživatelský interface a pozornost bude věnována i dalšímu ověřování kvality odezvy modelu a nastavení jeho parametrů pomocí dat z reálných sociálních sítí.

Reference

1. Ahmad, M. A. and Teredesai, A.: Modeling spread of ideas in online social networks. In: Proceedings of the *Fifth Australasian Conference on Data Mining and Analytics - Volume 61* (Sydney, Australia, November 29 - 30, 2006). P. Christen, P. J. Kennedy, J. Li, S. J. Simoff, and G. J. Williams, Eds. Conferences in Research and Practice in Information Technology Series, vol. 245. Australian Computer Society, Darlinghurst, Australia, 185-190, 2006.
2. Berger-Wolf, T. Y. and Saia, J.: A framework for analysis of dynamic social networks. In: Proceedings of the *12th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining* (Philadelphia, PA, USA, August 20 - 23, 2006). KDD '06. ACM, New York, NY, 523-528, 2006.
3. Borshchev A., Filippov A.: From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. *The 22nd International Conference of the System Dynamics Society*, Oxford, England, 2004.

4. Cha, M., Mislove, A., and Gummadi, K. P.: A measurement-driven analysis of information propagation in the flickr social network. In: *Proceedings of the 18th international Conference on World Wide Web* (Madrid, Spain, April 20 - 24, 2009). WWW '09. ACM, New York, NY, 721-730, 2009.
5. Forrester, J. W.: *Industrial dynamics*. Cambridge, MA: MIT Press, 1961.
6. Menges, F., Mishra, B., and Narzisi, G.: Modeling and simulation of e-mail social networks: a new stochastic agent-based approach. In: *Proceedings of the 40th Conference on Winter Simulation* (Miami, Florida, December 07 - 10, 2008). S. Mason, R. Hill, L. Mönch, and O. Rose, Eds. Winter Simulation Conference. Winter Simulation Conference, 2792-2800, 2008.
7. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B.: Measurement and analysis of online social networks. In: *Proceedings of the 7th ACM SIGCOMM Conference on internet Measurement* (San Diego, California, USA, October 24 - 26, 2007). IMC '07. ACM, New York, NY, 29-42, 2007.
8. Robinson, S., Alifantis, T., Hurriion, R., Ladbrook, J., Edwards, J., and Waller, T.: Modelling and improving human decision making with simulation. In: *Proceedings of the 33rd Conference on Winter Simulation* (Arlington, Virginia, December 09 - 12, 2001). Winter Simulation Conference. IEEE Computer Society, Washington, DC, 913-920, 2001..
9. Siebers, P.O. and Aickelin, U.: Introduction to Multi-Agent Simulation. In: *Encyclopedia of Decision Making and Decision Support Technologies*. IDEAS Group, 2007
10. Song, X., Lin, C., Tseng, B. L., and Sun, M.: Modeling and predicting personal information dissemination behavior. In: *Proceedings of the Eleventh ACM SIGKDD international Conference on Knowledge Discovery in Data Mining* (Chicago, Illinois, USA, August 21 - 24, 2005). KDD '05. ACM, New York, NY, 479-488, 2005.
11. Walter, F. E., Battiston, S., and Schweitzer, F.: Personalised and dynamic trust in social networks. In: *Proceedings of the Third ACM Conference on Recommender Systems* (New York, New York, USA, October 23 - 25, 2009). RecSys '09. ACM, New York, NY, 197-204, 2009.
12. Wasserman, S. and Faust, K.: *Social Network Analysis*. Cambridge University Press, 1994, ISBN 0521382696

Annotation:

Modeling dynamics of information flows in social networks

Social networks are currently of the most studied structures in the area of information and knowledge exchange due to its popularity among users. These structures are well described in terms of their static behavior; this paper tries to describe the dynamics as well as dissemination of information and knowledge in these networks. New heuristic model of individual behavior in a network is introduced here, based on individual's need for information and knowledge to solve specific situations. On that is constructed agent-based model of the whole social network, which was practically implemented and it's response on selected tasks will also be presented.