



Raport naukowo-techniczny  
Naukowa i Akademicka Sieć Komputerowa

# Wybrane podejścia do modelowania dynamiki sieci społecznych

Mariusz Kamola

Pracownia Sterowania Siecią, NASK

[mariusz.kamola@nask.pl](mailto:mariusz.kamola@nask.pl)

Raport nr: 2015-20

Warszawa, 31 grudnia 2015 r.

Copyright © 2015 by Naukowa i Akademicka Sieć Komputerowa.

Fragmenty tej publikacji mogą być kopiowane i cytowane pod warunkiem zachowania tekstu niniejszych zastrzeżeń w każdej kopii oraz powiadomienia Naukowej i Akademickiej Sieci Komputerowej.

**Streszczenie.** Dokument stanowi przegląd metod analizy i modelowania dynamiki sieci społecznych. Wprowadzany jest zgrubny podział zagadnień na związane z dynamiką atrybutów węzłów lub powiązań w sieciach o ustalonej strukturze oraz na związane z dynamiką samej topologii sieci: procesu powoływania nowych węzłów i wiązania ich z istniejącymi. Wyboru przedstawionych podejść dokonano mając na uwadze ich podobieństwo do klasycznych pojęć z dziedziny automatyki: macierzy stanu, stabilności, zmiennych sterujących. Warto podkreślić, że analiza cech sieci bazująca na wartościach i wektorach własnych macierzy sąsiedztwa jest równie użytecznym narzędziem, co odpowiadająca jej analiza stabilności układów dynamicznych. Różnice polegają na wymiarowości obu problemów i, wynikającej z niej – ograniczonej obserwowalności (w sensie potocznym) wielkich sieci społecznych. Powoduje to konieczność opracowywania strumieniowych metod analizy zdarzeń, nie mających, według wiedzy autora, swojego odpowiednika w automatyce.

## 1. Wstęp

Powszechność i różnorodność środków komunikacji powodują, że podejmowanie decyzji przez użytkownika w coraz większym zakresie jest przeprowadzane pod wpływem otoczenia. Otoczenie to stanowią inni użytkownicy mechanizmów komunikacji, przy czym użytkownik może na topologię tego otoczenia wpływać w specyficzny, ograniczony sposób.

Niniejsze opracowanie przedstawia wybrane, ważne i aktualne zjawiska dotyczące dynamiki zjawisk obserwowanych w systemach o strukturze sieci, w szczególności w sieciach społecznych. Dynamika ta zasadniczo może dotyczyć atrybutów węzłów w sieci o ustalonej strukturze; może też dotyczyć samej struktury sieci. Powody indywidualnych decyzji prowadzących do zmiany atrybutów lub struktury sieci często są właściwie takie same; jednak zagadnienia te rozpatrywane są w literaturze rozłącznie i tak też będzie tutaj. Zmiany atrybutów węzłów zostały przedstawione w rozdz. 2, a zmiany struktury – w rozdz. 3.

Uleganie przez użytkowników opinii otoczenia, czy wybór otoczenia wydawać się mogą spontaniczne, nieracjonalne – ale w rzeczywistości takie nie są. Wynikają one z, czasami nieświadomego, rachunku zysków i kosztów związanych z każdą decyzją. Zyski i koszty nie zawsze mają wymiar materialny, ale ostatecznie przedstawiają one konkretną wartość dla użytkownika, np. wpływając na jego komfort, samoocenę, wizerunek, perspektywy, bezpieczeństwo itp. Indywidualna hierarchia tych wartości różnicuje zachowania użytkowników w konkretnej sytuacji i kontekście działania (np. tematyce i sposobie funkcjonowania portalu społecznościowego czy też charakterze otoczenia, w którym aktualnie przychodzi mu działać). Podatność użytkownika na otoczenie jest zawsze ściśle związana z bilansem informacji będących w jego posiadaniu, w porównaniu z informacjami posiadanymi przez otoczenie – przy czym obie wartości są oceniane w sposób absolutnie subiektywny przez użytkownika.

## 2. Dynamika atrybutów węzłów sieci

### 2.1. Kaskady opinii

Niski poziom wiedzy użytkownika o otoczeniu czyni go podatnym na decyzje podejmowane przez innych. Książka [1], w rozdz. 16 poświęconym kaskadowemu rozprzestrzenianiu się opinii w sieci, posługuje się przykładem turysty w obcym mieście, zamierzającemu skorzystać z upatrzonej wcześniej restauracji. W sytuacji, gdy widzi on, że restauracja ta jest pusta, natomiast sąsiednia – pełna gości – decyduje się, paradoksalnie wbrew sobie, na tę ostatnią. Jest to decyzja racjonalna: turysta zwyczajnie nisko ocenił wartość posiadanych przez siebie informacji. W takich sytuacjach trend zapoczątkowany przez nawet bardzo niewielką liczbę użytkowników może zdominować zachowanie pozostałych. Konsekwentnie, wystarczy niewielka liczba użytkowników prezentujących odmienną opinię, aby rozwiać iluzję – lub zapoczątkować inną. Widać więc, że tzw. instynkt stadny ma swoje źródło w niedoinformowaniu użytkowników.

W sytuacji przedstawionej powyżej decyzja użytkownika była podyktowana bilansem informacyjnym (*informational effect*). Decyzje mogą być również powodowane przez prosty rachunek ekonomiczny zysków lub strat (*direct-benefit effect*). Taki schemat działania omówiony jest w [1], w rozdz. 19: poczynając od zdefiniowania interakcji użytkowników jako sieciowej gry z koordynacją, z macierzą wypłat użytkowników  $v$  i  $w$  za podjęcie decyzji  $A$  albo  $B$ :

|          |          |             |             |
|----------|----------|-------------|-------------|
|          |          | <i>w</i>    |             |
|          |          | <i>A</i>    | <i>B</i>    |
| <i>v</i> | <i>A</i> | <i>a, a</i> | <i>0, 0</i> |
|          | <i>B</i> | <i>0, 0</i> | <i>b, b</i> |

Decyzje *A* oraz *B* mogą być związane z wyborem jednej z dwóch konkurujących i wykluczających się technologii, partii politycznych, grona znajomych etc. i obowiązują wobec wszystkich sąsiadów konkretnego użytkownika. Proste kryterium ekonomiczne, bazujące na proporcji liczby sąsiadów, którzy wybrali *A* lub *B* oraz proporcji wypłat *a* lub *b* przesądza o wyborach poszczególnych użytkowników.

Zakres propagacji zmian atrybutów węzłów jest zdeterminowany strukturą samej sieci. Grona ściśle powiązanych użytkowników przyjmują taką samą wartość atrybutu i jest to stan stabilny. Mosty pomiędzy gronami mają niewystarczające oddziaływanie, aby skutecznie propagować wybór dokonany przez konkretne grono. Nie ma tu znaczenia rozmiar grona, a jedynie gęstość powiązań pomiędzy gronami oraz w gronach. Jedynie zmiana warunków gry, tj. wypłat, jest w stanie spowodować dalszą ekspansję wyboru *A* albo *B*.

## 2.2. Sterowanie opiniami

Zauważmy, że kaskady mogą być w pewnych okolicznościach inicjowane w sposób naturalny, determinowany przez strukturę grafu i charakter węzłów – inicjacja i przebieg kaskady przypomina wówczas nieco efekt motyla – mogą być one również inicjowane przez węzły o szczególnej sile, czy też przewadze nad innymi. W odniesieniu do sieci społecznych, mogą to być np. osoby o szczególnej charyzmie (autorytety moralne, modowe, kulturowe), zasobach finansowych (przekupujące wyborców, rozdające pierwsze egzemplarze noworodzącej się technologii) lub siłą podporządkowujące sobie i kształtujące nowe otoczenie. Naukowcy starają się modelować mechanizmy sterowania siecią głównie z wykorzystaniem klasycznej teorii sterowania. Np. artykuł [2] przedstawia dynamikę sieci, w której stan węzłów jest analogiczny ze zmienną stanu w automatyce. Konsekwentnie, macierz sąsiedztwa jest tożsama z macierzą stanu układu dynamicznego. Takie podejście modeluje dość specyficzną interakcję węzłów, gdzie atrybut węzła jest ciągły, skalarny, a jego pochodna jest proporcjonalna do wynikowego oddziaływania nań przez węzły sąsiednie.

W takim grafie wyróżnia się węzły sterujące, niepoddające się wpływowi innych, działające wg szczególnych zasad. Odpowiadają one wejściom sterującym układów dynamicznych z jednej strony; z drugiej strony – jednostkom o szczególnej sile, opisanym powyżej. Węzły sterujące mogą powstawać w sieci w sposób naturalny, ale również możliwe jest kształtowanie niektórych węzłów do tej roli przez regulatora (np. przez selektywne wsparcie finansowe lub inne) – wówczas zagadnienie optymalnego wyboru węzłów sterujących odpowiada znanemu w automatyce od lat zagadnieniu wyboru struktury sterowania. Co ciekawe, wykazano na przykładzie znanych sieci [2], że wybór węzłów o wysokiej centralności nie zawsze jest optymalny.

Zauważmy, że czerpanie z bogatego arsenału środków optymalnego sterowania, jakie daje automatyka, jest możliwe w odniesieniu do sieci o co prawda nieograniczonej złożoności topologicznej, ale również o dość szczególnie zdefiniowanym sposobie oddziaływania węzłów. Dynamika interakcji zakłada, że węzły nieustannie (a często również synchronicznie) podejmują decyzje; co więcej – i najważniejsze – zależne liniowo od poprzednich decyzji sąsiadów. Ta liniowość i brak antycypowania przyszłego stanu sieci stanowią oczywiście uproszczenie rzeczywistości, ale jednocześnie są dobrym punktem wyjścia do rozwoju bardziej skomplikowanych metod sterowania siecią. Zresztą, pokazano [3] liniową zależność punktów równowagi w graphach pomiędzy węzłami sieci od ak-

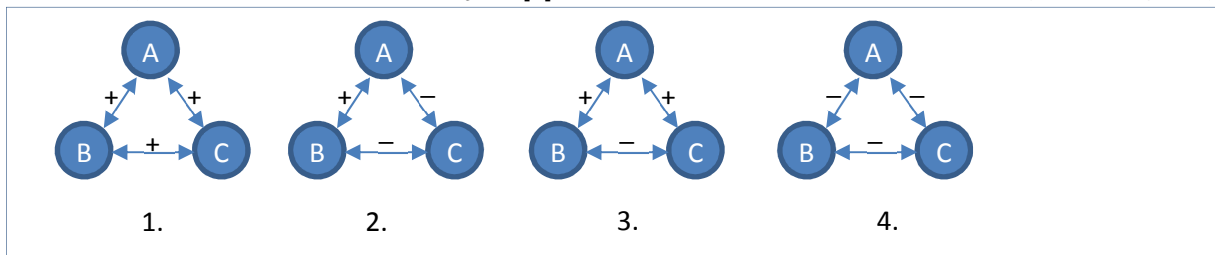
tualnych wartości „stanów” tych punktów, co sprowadza niektóre scenariusze teoriogrowe do problemu optymalnego sterowania. Wykorzystując w takich przypadkach klasyczne narzędzia automatyki (funkcja Lapunowa, równanie Riccatiego) można, dla założonych węzłów sterowania (wskazanych zygzakami na rys. obok) definiować typowe zadania sterowania optymalnego – czasooptymalne, z ograniczeniami stanu i sterowania itp.



Przedstawiona analogia sieci i układów dynamicznych jest uprawniona w szczególnych przypadkach, tj. głównie ciągłości parametru charakteryzującego węzły, iteracyjności i liniowości zależności decyzji podejmowanych przez jednostki w funkcji zachowań własnego, najbliższego otoczenia. Jednak jeśli rozważymy reakcje nieliniowe, wówczas podejście „automatyczne”, formalnie poprawne może być niewykonalne z uwagi na rozmiar sieci i przekleństwo wymiarowości, właściwe zadaniom programowania dynamicznego. W praktyce jak najbardziej obserwuje się w sieciach społecznych zarówno dyskretyzację podejmowanych decyzji, jak i różne kategorie reakcji użytkowników na bodźce otoczenia.

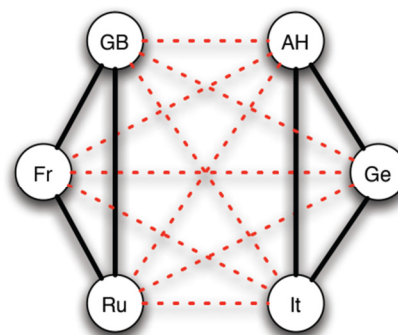
### 2.3. Triady i stabilność strukturalna

Znanym od dawna, podstawowym scenariuszem interakcji węzłów sieci, jest przypadek z polaryzacją oddziaływania: dwaj sąsiedzi mogą się albo lubić (+), albo nie (-). Analiza triad z tak określonymi oddziaływaniami jest prowadzona od lat 40. ub. wieku; dotychczasowe osiągnięcia zostały zebrane i jasno przedstawione w książce [1], w rozdz. 5. Spośród czterech możliwych konfiguracji



polaryzacji połączeń międzywęzłowych (por. rys. powyżej) tylko dwie pierwsze są stabilne. Konfiguracja 1. nie wymaga komentarza: wzajemne sympatie stabilizują trójkąt. W konfiguracji 2. węzły A i B sprzymierzają się przeciwko wspólnemu wrogowi, co również jest sytuacją trwałą. W konfiguracji 3. węzeł A będzie dążył do pogodzenia B i C, co w przypadku powodzenia doprowadzi do konfiguracji 1., inaczej wynikiem będzie konfiguracja 2. Najmniej oczywista (i najrzadziej obserwowana w praktyce) jest niestabilność konfiguracji 4. – zauważono, że skutkuje ona sprzymierzeniem się dwóch węzłów przeciwko trzeciemu.

Stabilność całej sieci można określać, analizując odsetek zauważonych w niej trójkątów stabilnych i niestabilnych. Frank Harary sformułował w 1953 r. twierdzenie o strukturze zrównoważonej sieci: graf pełny jest silnie zrównoważony strukturalnie, jeśli albo wszystkie węzły mają relacje przyjacielskie, albo można wyróżnić podzbiory  $X$  i  $Y$  węzłów takie, że w obrębie każdego z tych zbiorów występują wyłącznie relacje przyjacielskie, a relacje pomiędzy zbiorami  $X$  i  $Y$  są wyłącznie antagonistyczne. Znakomitą historyczną demon-



stracją twierdzenia jest układ sił pomiędzy mocarstwami europejskimi przed wybuchem I wojny światowej (rys. powyżej). Słabszą formą stabilności strukturalnej obserwujemy gdy możemy wyróżnić więcej podzbiorów analogicznych do  $X$  i  $Y$ , tj. z w pełni przyjacielskimi relacjami wewnątrz zbiorów i w pełni antagonistycznymi pomiędzy zbiorami.

## 2.4. Praktyczne modele reakcji

Zauważmy, że modele kaskad informacyjnych i stabilności strukturalnej dotyczą sytuacji z dyskretnym stanem węzłów sieci i w pewnym zakresie się uzupełniają. Pierwszy umożliwia wygodne modelowanie dynamiki sieci krok po kroku, drugi natomiast odzwierciedla zjawisko sprzymierzania się wcześniejszych wrogów. Natomiast modele liniowe wykorzystywane do sterowania sieci wymagają silnych założeń o jednolitym sposobie funkcjonowania węzłów. W praktyce węzły - odpowiadające użytkownikom - reagują w sposób niejednorodny na bodźce otoczenia. Wynika to niekiedy z nieracjonalności decyzji, a raczej ze zróżnicowanej oceny własnej wiedzy przez poszczególnych użytkowników. Obserwowane podstawowe reakcje zostały sklasyfikowane już ponad 60 lat temu [4]; są nimi:

- upór (*subbornness*)
- kompromis (*compromise*)
- konformizm stronniczy (*biased conformity*)

Użytkownicy uparci nie uwzględniają opinii otoczenia. Opinia  $z_i(t)$  użytkownika kompromisowego  $i$  w chwili  $t$  jest wypadkową opinii otoczenia, np. w modelu DeGroota [5]:

$$z_i(t+1) = \frac{w_{ii}z_i(t) + \sum_{j \in N(i)} w_{ij}z_j(t)}{w_{ii} + \sum_{j \in N(i)} w_{ij}}$$

gdzie  $w_{ij}$  to wagi krawędzi. Alternatywnie [6], można uwzględnić niezmienność własnej opinii, zastępując  $w_{ii}z_i(t)$  wartością stałą. Użytkownicy konformistyczni przyjmują w całości jedną z opinii swoich sąsiadów. Modele decyzyjne takich sytuacji zostały omówione w przypadku kaskad; postulowana stronniczość wynika z wag krawędzi, reprezentujących autorytet poszczególnych sąsiadów. W ogólności, modele zachowań stadnych określane są również modelami z głosowaniem, gdyż ostatecznie wybór użytkownika jest bezkompromisowy; wyklucza przyjęcie w części lub całości innej opinii. Nie dyskretność wyboru (por. *kupię sobie* i *peceta*, i *Maca*), ale bezkompromisowość (*kupię sobie* albo *peceta*, albo *Maca*) jest wyróżnikiem tego modelu.

Ponieważ w praktyce obserwowane są w jednej sieci społecznej wszystkie typy reakcji, proponowane są [7] modele łączone, w których węzły przyjmują jedną ze strategii z założonym prawdopodobieństwem. Dynamika tak modelowanych sieci jest zasadniczo zgodna z wynikami eksperymentalnymi z udziałem ochotników. Okazuje się ponadto, że na wybór reakcji ma wpływ wielkość sąsiedztwa; przy dużej liczbie sąsiadów preferowana jest reakcja konformistyczna. Przy małej - upór.

## 3. Dynamika struktury sieci

### 3.1. Klasyfikacja zadań

Zagadnienia dotyczące badania dynamiki struktury sieci obejmują zarówno tworzenie modeli matematycznych ewolucji struktury pojedynczych powiązań w sieci, jak też dynamiki zjawisk makroskopowych. Obejmują również opracowywanie algorytmów analizy zjawisk przejściowych w sieci -

widzianej jako całość, albo z perspektywy pewnej sekwencji zmian, obserwowanej w założonym okresie czasu. Artykuł przeglądowy [8] wprowadza zgrubną klasyfikację problemów analizy i modelowania ze względu na intensywność samych zmian oraz ze względu na zdolność wykrywania zmian strukturalnych.

Ze względu na intensywność zmian, zagadnienia analityczne dzielimy na:

- migawkowe (*slowly evolving*) – operujące na migawkach, tj. kompletnych grafach, obserwowanych w kolejnych chwilach pomiarowych; takie podejście jest możliwe dla grafów o niezbyt dużej dynamice i rozmiarze;
- zdarzeniowe (*streaming*) – reagujące na kolejne informacje o poszczególnych zmianach w sieci; właściwe dla dowolnie dużych grafów, ale z definicji nie przechowujące informacji o całym grafie.

Ze względu na zdolność wykrywania zmian strukturalnych, zagadnienia analityczne dzielimy na:

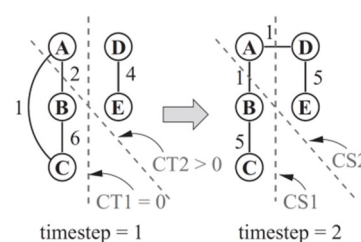
- aktualizujące wiedzę (*maintenance*) – uaktualniające posiadane już wyniki analityczne dotyczące struktury sieci (np. grup, średnic, mostów);
- wykrywające zmianę strukturalną (*evolution analysis*) – wykrywające zachodzące zmiany strukturalne i osobliwości w sieci (np. perkolację, anomalie).

Zagadnienia analizy i modelowania można zatem podzielić na cztery grupy, dla każdej kombinacji powyższych atrybutów. Tak też zostaną przedstawione, choć należy pamiętać, że zaproponowana w [8] klasyfikacja nie wyklucza istnienia zagadnień z pogranicza, albo wręcz przynależących od kilku klas naraz, np. płynnie przechodzących od analizy migawkowej do zdarzeniowej, wg jakiegoś parametru.

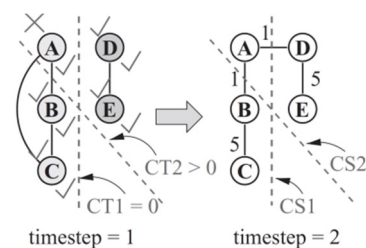
### 3.2. Aktualizacja posiadanej wiedzy – migawkowa

Migawkowe podejście przyrostowe polega na powtarzaniu analizy sieci dla poszczególnych migawek, ale w związku z wynikami poprzednich analiz tak, aby zachować tożsamość wcześniej odkrytych zjawisk. Typowe zadania analityczne związane są z grupowaniem i klasyfikacją węzłów.

Sekwencyjne **grupowanie węzłów** ma za zadanie – oprócz zachowania tradycyjnej dokładności analizy (*consistency*) – zachowanie podobieństwa wyników z kolejnych migawek (*smoothness*). Podobieństwo to może być identyfikowane dwojako [9]. Można więc unikać nowych podziałów na grupy, które zastosowane wobec danych historycznych skutkowałyby rażąco niekorzystnymi podziałami. Na rysunku obok przedstawiono dwa proponowane podziały dla nowej migawki (CS1, CS2, timestep=2). Drugi, rozpatrywany wyłącznie dla nowych danych, jest korzystniejszy, lecz w odniesieniu do danych historycznych (timestep=1) jest bardzo zły, bo przecina aż dwie krawędzie. Możliwe są różne kryteria ocen jakości podziału, jak również różne mechanizmy grupowania.

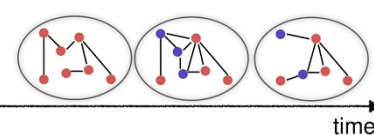


W innym podejściu nadrzędnym celem może być zachowanie przypisania węzłów do pierwotnych grup. Wówczas rozpatrujemy podziały grafu dla nowej migawki pod kątem wprowadzanych przez nie zmian przynależności węzłów do poprzednio stworzonych grup. Jeśli przyjmiemy, że macierz binarna  $Z_{n \times k}$  określa przynależność węzłów do grup, macierz  $Z_t^*$  jest unormowaną, ortonormalną wersją macierzy przynależności, dla migawki wykonanej w chwili  $t$ , to podobieństwo podziału w kolejnych migawkach określa macierz  $W$ , dobierana tak, by zminimalizować ślad:  $tr(Z_t^{*T} W_{t-1} Z_t^*)$ . Na rysunku obok przedstawiono analogiczne podziały grafu: tutaj również podział CS2 okazuje się gorszy, ale dlatego, że zmienia przypisanie węzła A do grupy.



### 3.3. Aktualizacja posiadanej wiedzy – zdarzeniowa

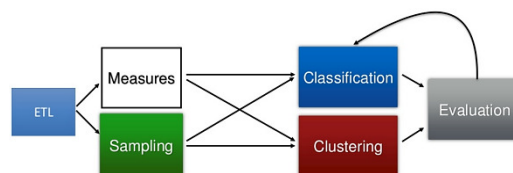
W podejściu zdarzeniowym zakładamy, że sieć  $G$  jest zbyt duża, aby możliwe było lub opłacało się ją analizować w całości. Zamiast tego obserwujemy strumień zdarzeń – grafów  $G_i$  – opisujących przyrostowo rozwój sieci  $G$  (por. rys. obok). Typowe zadania analityczne obejmują, podobnie jak poprzednio, estymację istniejących grup w grafie oraz klasyfikację węzłów wyłącznie na podstawie danych przyrostowych, tj. obserwacji strumienia  $G_1, G_2, \dots, G_i$ .



Jedną z interesujących koncepcji [14] rozwiązania problemu przyrostowego **grupowania węzłów** polega na wstępnym wyznaczeniu zbiorów  $V_j$  węzłów stosunkowo często współwystępujących w obserwowanym strumieniu. Skłonność do współwystępowania (*affinity*) dla arbitralnego zbioru  $V_j$  definiowana jest jako iloraz częstości występowania grafów zawierających *wszystkie* wierzchołki z  $V_j$  i częstości występowania grafów zawierających *co najmniej jeden* wierzchołek z  $V_j$ . Sens tego współczynnika jest podobny do wskaźnika *tf-idf* w wyszukiwaniu pełnotekstowym – wyższą rangę uzyskują słowa (lub zbiory słów) niepospolite.

W obrębie zbiorów  $V_j$  o najwyższym rankingu wyszukiwane są następnie te, które charakteryzują się dużą średnią gęstością podgrafów na nich rozpiętych, obliczoną tylko dla grafów ze strumienia, które zawierają *wszystkie* węzły z  $V_j$ . Tak sformułowane zadanie wyszukiwania *pseudoklik* musi podczas implementacji zostać zoptymalizowane dla uzyskania zadowalającej wydajności - stosowane są wówczas m.in. funkcje skrótu.

Analiza zdarzeniowa doczekała się już implementacji w postaci ogólnodostępnych narzędzi, takich jak Apache Flink, wykonujących typowe zadania: grupowania i klasyfikacji, również na dynamicznie łączonych strumieniach danych z wielu źródeł (por. rys. obok.).



### 3.4. Wykrywanie zmian strukturalnych – migawkowe

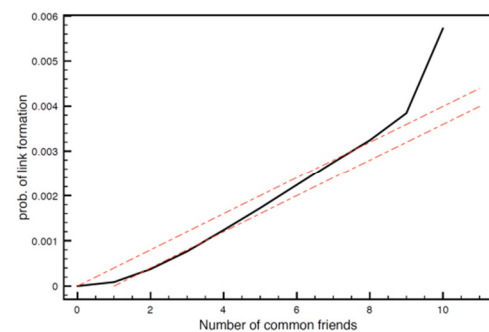
Analiza struktury sieci ma na celu wykrycie oraz opis mechanizmu zachodzenia istotnych zmian strukturalnych, tj. takich, które determinują sposób funkcjonowania całej sieci. Mogą być nimi: powstanie dominującej składowej spójnej, raptowna zmiana średnicy (albo ogólniej, struktury grafu) czy perkolacja. Droga do wychwycenia takich osobliwości prowadzi jednak przez mozolne modelowanie zachowania węzłów w sieci tak, aby dobrze odzwierciedlić jej rozwój w kolejnych chwi-



lach czasowych-migawkach. Dlatego podstawowym podejściem jest modelowanie dynamiki krawędzi w grafie.

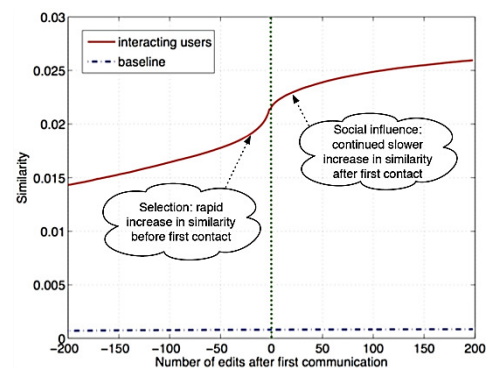
Modelowanie **powstawania i zanikania** krawędzi jest obecnie bardzo rozwiniętym działem badań. Najbardziej znany, podstawowy model wzrostu sieci opracowali A.-L. Barabasi i R. Albert [15]: każdy nowotworzony węzeł tworzy stałą liczbę połączeń z już istniejącymi, przy czym prawdopodobieństwo dołączenia się do istniejącego węzła zależy od stopnia tego węzła. Takie preferencyjne dołączanie tworzy sieć o rozkładzie potęgowym, lecz z małym współczynnikiem gronowania oraz zerową asortatywnością. Wprowadzenie nieliniowego preferencyjnego dołączania umożliwia tworzenie sieci dysasortatywnych. Innym klasycznym modelem zaproponowali P. Erdős i A. Rényi [16]; polega on na losowym łączeniu istniejących, niepołączonych wcześniej węzłów. Prowadzi on do powstania sieci małych światów, ale nie – bezskalowej.

Dalsze badania umożliwiły odkrycie kolejnych czynników determinujących prawdopodobieństwo powstawania krawędzi w grafie. Analiza korespondencji mailowej potwierdziła istotną rolę homofilii w nawiązywaniu połączeń, wg zasady: moi dwaj nieznanymi się przyjaciele prędzej czy później poznają się i zostaną przyjaciółmi. Okazuje się [13], że prawdopodobieństwo powstania krawędzi łączącej dwa węzły jest niemalże proporcjonalne do liczby wspólnych sąsiadów tych węzłów, por. wykres obok.



Zjawisko to zostało wykorzystane w innym, bardzo popularnym modelu wzrostu sieci [17], zakładającym dołączanie nowego węzła do *pary* istniejących i połączonych już krawędzią.

Należy jednak pamiętać, że wspólni znajomi nie muszą być sami w sobie powodem powstania połączenia, a jedynie wskazują na istnienie wspólnych cech łączących się węzłów. Tak zaobserwowano w przypadku twórców haseł w Wikipedii [12]: wielu z nich nawiązuje w końcu bezpośredni kontakt mailowy, ale powodem tego jest wcześniejsze, niezwiązane z jakimikolwiek zachowaniami społecznymi, wspólne zainteresowanie pewną tematyką. Późniejszy, coraz silniejszy związek między twórcami, rozwija się już głównie na bazie bezpośredniego związku międzyludzkiego, por. wykres obok.

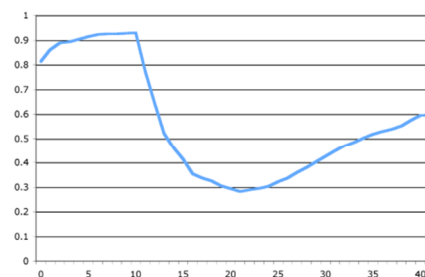


Wnikliwe obserwacje dynamiki i struktury powiązań w portalach społecznościowych *Yahoo!* i *Flickr* [10] zaowocowały rozwinięciem idei preferencyjnego dołączania, uwzględniającym trzy zdefiniowane role społeczne użytkowników portali społecznościowych:

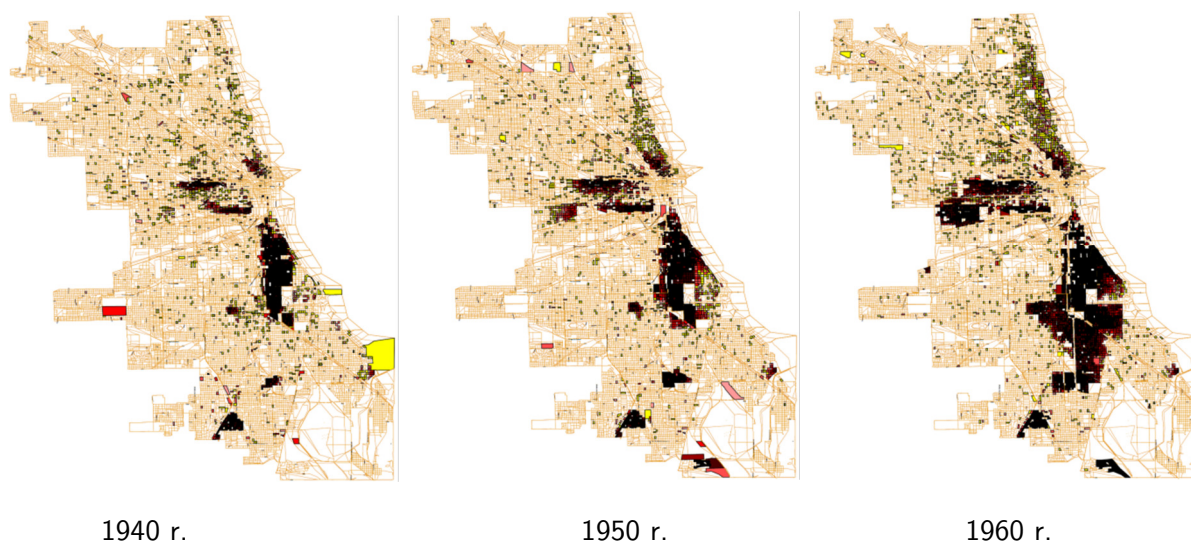
- zapraszających – *inviters* – do portalu ludzi dotychczas z nim niezwiązanych,
- dusze towarzystwa – *linkers* – stymulujących wzrost liczby powiązań,
- pasywnych – *passive* – czekających na zaproszenie.

Model generuje węzły trzech powyższych typów; ich rola jest później niezmienna. Role aktywne dokonują okresowego dołączania do siebie lub wzajemnego łączenia węzłów odpowiednich typów. Struktura sieci wygenerowanej w ten sposób ewoluje dość zgodnie z zaobserwowanymi faktycznymi sieciami społecznymi, wykazując powstanie dominującej składowej spójnej, izolowanych spo-

łeczności oraz węzłów niepołączonych. Dodatkowo, w obu przypadkach rzeczywistych zaobserwowano dość szczególną ewolucję gęstości grafu, w postaci następujących faz: euforycznej, wyciszenia i organicznego wzrostu, por. wykres obok.



Szczególnym scenariuszem dynamiki struktury sieci jest przypadek, w którym z każdym węzłem może (lecz nie musi) być związany co najwyżej jeden identyfikator, reprezentujący jednostkę podejmującą autonomiczne decyzje, tzw. *agenta* o określonym typie. W scenariuszu agent może migrować do wolnych węzłów w grafie o ustalonej strukturze. Nie zmienia się więc sama struktura dopuszczalnych powiązań, ale zmienia się struktura powiązań faktycznych, gdyż jedynym motywem migracji jest chęć zasiedlenia takiego węzła, aby w bezpośrednim sąsiedztwie znajdowało się co najmniej  $k$  węzłów o tym samym typie, co migrujący agent. Jest to tzw. **model Schellinga**; opisuje od dobrze zachowania stadne, jak np. proces segregacji rasowej osadnictwa czarnoskórych w Chicago [11], por. mapy poniżej. Zachowania takie są spontaniczne; nie wymagają one żadnej stymulacji zewnętrznej. Wielkość powstających, stabilnych społeczności zależy od owego  $k$ , pragnienia grupowania się. Szybkość ustalania się struktury grafu zależy również od nadwyżki wolnego miejsca w grafie.



Do przyrostowej **analizy struktury** sieci wykorzystuje się metody widmowe, czyli bazujące na rozkładzie macierzy sąsiedztwa  $A$  na iloczyn macierzy ortonormalnej  $P$  (obrotu przestrzeni) i macierzy diagonalnej wartości własnych  $D$  (rozciągnięcia przestrzeni – jeśli interpretujemy  $A$  jako macierz odwzorowania):  $A = PDP^T$ . Wektor własny odpowiadający najistotniejszej, w sensie modułu, wartości własnej, wskazuje „główne kierunki przepływu informacji w sieci”. Istotna tego kierunku, rozumiana jako kąt pomiędzy najważniejszymi wektorami własnymi w kolejnych migawkach, informuje o znaczącej zmianie w strukturze sieci [18]. Inne, ciekawe podejście [19] wprowadza model autoregresyjny ewolucji grafu, aby wychwytywać anomalie.

### 3.5. Wykrywanie zmian strukturalnych – zdarzeniowe

Zachodzenie zmian strukturalnych – tak w analizie migawkowej, jak i zdarzeniowej – jest utożsamiane z **wykrywaniem anomalii** (*outlier detection*), gdyż każda taka zmiana w sposób istotny wpływa na parametry całej sieci, np. początek współpracy naukowej badaczy z dwóch, dotąd niezwiązanych, dziedzin zmniejsza średnicę grafu i zmienia centralność węzłów odpowiadających tym

badaczom. O ile analiza migawkowa pozwala na wgląd w całą strukturę grafu, w podejściu przyrostowym mamy możliwość wnioskowania tylko na podstawie danych szczątkowych. Można powiedzieć, że strumień obserwowanych zdarzeń jest pewnym rodzajem próbkowania statystycznego. Wnioskowanie o globalnych zmianach strukturalnych jest więc obarczone większym błędem. Niemniej jednak wśród propozycji możemy znaleźć pomysły wnioskowania w oparciu o wykryte, nieobserwowane wcześniej powiązania pomiędzy zaewidencjonowanymi grupami węzłów [20]; podejście to jest logicznym uzupełnieniem zdarzeniowego grupowania: jeśli mamy już ustaloną wiedzę o grupach, to nagłe, nieefemeryczne pojawienie się połączenia między grupami jest zdarzeniem doniosłym.

Inne, ciekawe podejście [21] skupia się wokół monitorowania otoczenia każdego z węzłów – rozpatrywanego jako podgraf, nieznaney z definicji, całej sieci. Parametry podgrafów wynikają z analizy widmowej – są to wartości i wektory własne. Wektory własne interpretowane są jako wzorce aktywności poszczególnych węzłów (np. grono węzłów, z którymi zachodzi interakcja), podczas gdy odpowiadające im wartości własne opisują intensywność oddziaływania. Monitorowanie zmian obu tych parametrów umożliwia wychwycenie zmian w zachowaniu poszczególnych węzłów.

## Bibliografia

- [1] Easley, David, and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [2] Liu, Yang-Yu, Jean-Jacques Slotine, and Albert-László Barabási. "Controllability of complex networks." *Nature* 473.7346 (2011): 167-173.
- [3] Bloembergen, Daan, et al. "Influencing social networks: An optimal control study." *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI'14)*. 2014.
- [4] Asch, Solomon E. "Opinions and social pressure." *Readings about the social animal* 193 (1955): 17-26.
- [5] DeGroot, Morris H. "Reaching a consensus." *Journal of the American Statistical Association* 69.345 (1974): 118-121.
- [6] Friedkin, Noah E., and Eugene C. Johnsen. "Social influence and opinions." *Journal of Mathematical Sociology* 15.3-4 (1990): 193-206.
- [7] Das, Abhimanyu, Sreenivas Gollapudi, and Kamesh Munagala. "Modeling opinion dynamics in social networks." *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 2014.
- [8] Aggarwal, Charu, and Karthik Subbian. "Evolutionary network analysis: A survey." *ACM Computing Surveys (CSUR)* 47.1 (2014): 10.
- [9] Chi, Yun, et al. "On evolutionary spectral clustering." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3.4 (2009): 17.
- [10] Kumar, Ravi, Jasmine Novak, and Andrew Tomkins. "Structure and evolution of online social networks." *Link mining: models, algorithms, and applications*. Springer New York, 2010. 337-357.
- [11] Möbius, Markus M., and Tanya S. Rosenblat. "The process of ghetto formation: Evidence from Chicago." Unpublished manuscript. <http://trosenblat.nber.org/papers/Files/Chicago/chicago-dec7.pdf> (2001).
- [12] Crandall, David, et al. "Feedback effects between similarity and social influence in online communities." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.
- [13] Kossinets, Gueorgi, and Duncan J. Watts. "Empirical analysis of an evolving social network." *Science* 311.5757 (2006): 88-90.
- [14] Aggarwal, Charu C., et al. "On dense pattern mining in graph streams." *Proceedings of the VLDB Endowment* 3.1-2 (2010): 975-984.

- [15] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *Science* 286.5439 (1999): 509-512.
- [16] Erdős, P.; Rényi, A. "On Random Graphs. I". *Publicationes Mathematicae* 6: 290–297 (1959)
- [17] Dorogovtsev, S. N. and Mendes, J. F. F., „Evolution of networks”, *Adv. Phys.*, 2002, 1079–1187
- [18] Ide, Tsuyoshi, and Hisashi Kashima. "Eigenspace-based anomaly detection in computer systems." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
- [19] Shoubridge, Peter, et al. "Detection of abnormal change in a time series of graphs." *Journal of Interconnection Networks* 3.01n02 (2002): 85-101.
- [20] Aggarwal, Charu C., Yuchen Zhao, and Philip S. Yu. "Outlier detection in graph streams." *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 2011.
- [21] Yu, Weiren, et al. "On anomalous hotspot discovery in graph streams." *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013.