

Andrzej Jarynowski

Stephanie Boland

Analiza sieci społecznych w odkrywaniu narracyjnej struktury fikcji literackiej

1) Wprowadzenie

1a) Ilościowe badania tekstów

W związku z rozwojem metod komputerowych w przetwarzaniu informacji pojawiło się pole do wykorzystania automatów w analizie literackiej. Proste czynności zliczeniowe, które dało się zautomatyzować, zostały adoptowane do warsztatu badacza literatury. Jednak prawdziwy przełom został spowodowany wykorzystaniem narzędzi, ukazujących korelację między różnymi jednostkami analizy. Umożliwiły to techniki przetwarzania języka naturalnego (ang. *natural language processing* - NLP), które pozwoliły analizować treści strukturyzowane składniowo i gramatycznie w ramach określonego kodu językowego. Metody i techniki NLP zostały opracowane przez informatyków w celu ułatwienia przetwarzania informacji zapisanych w sposób zrozumiały dla człowieka, a nie dla systemów analitycznych. Dopiero przetransformowanie takich informacji pozwala analitykom na zautomatyzowanie swojej pracy. Początkowo techniki NLP służyły dokonywaniu analizy niestrukturyzowanych danych w postaci zwykłego tekstu. Na przykład bez konieczności czytania tekstów dokonywano oceny nacechowania emocjonalnego wypowiedzi na (podstawie) doboru słów oraz ich struktury. Techniki opisane powyżej znalazły wiele komercyjnych zastosowań i są coraz silniej rozwijane w celu jak najdokładniejszej klasyfikacji tekstów.

Nie jest to jednak ich jedyne zastosowanie, gdyż badania ilościowe w literaturze zagościły już na dobre w literaturoznawstwie czy językoznawstwie. Najlepszym przykładem takiego podejścia jest analiza Zipfa¹, gdzie (w której) bada się rozkład częstości występowania słów w tekście. Okazuje się, że na podstawie obserwacji częstości występowania słów można odróżniać pisarzy czy gatunki literackie. Ostatnio nawet w Instytucie Fizyki Jądrowej w Krakowie² analizowane są dodatkowo połączenia między słowami (ich umiejscowienie obok siebie w zdaniu), które można przedstawić w postaci sieci.

1b) Analiza sieciowa

W naszej pracy właśnie analiza sieciowa jest podstawowym narzędziem metodologicznym. Teoria sieci złożonych niezwykle dynamicznie rozwija się od publikacji Barabásiego³ o grafach przypadkowych. Wielką przewagą podejścia sieciowego jest ogromne pole potencjalnych zastosowań od układów fizycznych, przez biologiczne, po społeczne. Można ich używać właściwie

1 GK. Zipf, *The Psychobiology of Language*, Houghton-Mifflin 1935.

2 A. Kulig, *Literary and scientific texts in network representation*, w:6 Ogólnopolskie Sympozjum "Fizyka w Ekonomii i Naukach Społecznych", Gdańsk 2012, prezentacja ustna

3 R. Albert, AL. Barabási, *Statistical mechanics of complex networks*. Rev. Mod. Phys. 74: 47–97 2002

wszędzie, gdzie istnieją zależności między elementami. W międzyczasie na pograniczu nauk ścisłych oraz społecznych wyodrębniła się nowa technika badawcza: analiza sieci społecznych (ang. *social network analysis*- SNA). Struktura powiązań między ludźmi decyduje o wielu czynnikach takich, jak przepływ informacji lub stosunek władzy. Położenie jednostki w sieci również warunkuje wiele cech, jak choćby pozycja społeczna. W ramach teorii SNA wyróżniamy sieci binarne (połączenie między elementami istnieje, bądź nie) oraz ważone (połączenia mogą mieć różne wagi, odzwierciedlające różne poziomy intensywności interakcji). W naszym przypadku intensywność relacji będzie miała znaczenie w celu dokładniejszego opisu. Ponadto znamy sieci skierowane (połączenie między elementami ma ustalony kierunek) oraz nieskierowane (połączenia są bez kierunku, a notowana jest tylko relacja wzajemności).

1c) Percepcja sieci w literaturze

Wiedząc, jak ważne są sieci społeczne w rzeczywistym świecie, warto podjąć się oceny roli takiej sieci między bohaterami w świecie fikcyjnym. Najpierw należy zatrzymać się na samym procesie percepcji sieci oczami czytelnika postawionego przed konstruktorem narracyjnym utworzonym przez autora. Niniejsza praca dotyczy właśnie tematu postrzegania sieci w różnych kontekstach czytelniczych w porównaniu z komputerowymi metodami automatycznej detekcji. Prowadzone przez nas rozważania są badaniami eksploracyjnymi. Wedle naszej wiedzy, nikt wcześniej nie analizował sieci społecznych w literaturze pod kątem percepcji. Dlatego też, skoncentrowaliśmy się na interpretacji sieci społecznych, odczytywanych subiektywnie przez czytelnika, napominając tylko o innych kontekstach, które w przyszłości powinny zostać przebadane.

1d) Ekstrakcja sieci

Warto jednak umiejscowić nasze badania w szerszym kontekście dotychczasowej wiedzy na temat sieci społecznych w literaturze. Prym w tej dziedzinie wiodą trzy amerykańskie ośrodki: grupa NLP z Uniwersytetu Columbia w Nowym Jorku, „Literary Lab” z Uniwersytetu Stanford w Kaliforni, oraz grupa „Google Books” z Uniwersytetu Harvarda w Bostonie. Tego typu badania zaliczają się do cyfrowej humanistyki (ang. *digital humanities*). Każdy z tych zespołów skupił się na trochę innym aspekcie, niemniej jednak wspólnym mianownikiem było skoncentrowanie się na automatycznej ekstrakcji sieci. Pierwsze prace dotyczyły analizy utworów dramatycznych, gdzie połączeniem między bohaterami (postaciami) była ich obecność w tej samej scenie. Franco Moretti ze Stanford, koordynujący te działania, pokazał właściwości tych sieci na przykładzie dzieł Szekspira⁴. Później porównał je z dramatami chińskimi. Jedną z najważniejszych poczynionych obserwacji była znacząca przewaga w ilości połączeń społecznych (liczba znajomych) centralnych postaci z dzieł europejskich w porównaniu z chińskimi, gdzie postaci miały bardziej równomiernie rozłożone znajomości po społeczności. Naukowcy z Harvardu, wśród których jest Martin Nowak⁵,

4 F. Moretti, *Network Theory, Plot Analysis*, New Left Review 68, March-April 2011.

5 JM. Babbitt et al, *Quantitative analysis of culture using millions of digitized books*, Science 331 (6014) 2011.

szukali asocjacji pomiędzy różnymi hasłami, pojawiającymi się w zbiorze książek zdigitalizowanych przez Google. Jednak punktem wyjścia do naszej analizy były prace grupy z Uniwersytetu Columbia, w którym m.in. wyekstrahowano sieci kontaktów konwersatoryjnych bohaterów dziewiętnastowiecznych powieści wiktoriańskich⁶. W tym celu skonstruowano algorytm oparty na liczeniu maszynowym, który na początku ustala listę bohaterów. W związku z przyjętą definicją sieci kontaktów poprzez sieć konwersatoryjną uznawane są sytuacje dialogowe w postaci rozmowy oraz tekstu akapitowego w formie mowy niezależnej, jak i zależnej. Tak utworzone sieci pokazywały np. zasadniczą różnicę między pierwszo- a trzecioosobową narracją. Uznaliśmy jednak, iż zawężenie interakcji społecznych do konwersacji, bardzo sensowne z punktu widzenia automatycznej ekstrakcji, niestety wymaga uzupełnienia i weryfikacji poprzez porównanie z wizją czytelników. Tak wąsko rozumiane sieci korelowały tylko z wynikami kwestionariuszy, wypełnionych przez osoby, którym zadano wytyczne zaznaczania interakcji zgodnie z algorytmem automatycznej anotacji.

2) Wybór materiału i konceptualizacja badania

Znając dotychczasowy stan wiedzy w dziedzinie cyfrowej humanistyki wybraliśmy praktycznie nieporuszony temat percepcji sieci jako takiej. W tym celu postanowiliśmy przeprowadzić eksperyment, w którym porównywane byłyby komputerowe algorytmy ekstrakcji sieci z subiektywną percepcją uczestników badań. Samo zróżnicowanie poziomu czytelników i ich opinii było również istotne dla badania, a nawet najważniejsze z perspektywy tego artykułu. Do pracy nad tekstem wybraliśmy tom opowiadań Sherwooda Andersona pod tytułem „Winesburg, Ohio” należący do kanonu literatury anglosaskiej. Wybór na to dzieło padł z dwóch powodów. Po pierwsze, mamy tu do czynienia z opisem życia obyczajowego społeczności małego miasteczka. To właśnie interakcje społeczne budują fabułę opowieści (co próbowaliśmy później weryfikować ilościowo). W przypadku tego autora warto wspomnieć także o dużej ilości bohaterów, którzy tworzą zamknięty zbiór postaci, pojawiających się w całym zbiorze jego opowiadań⁷. Po drugie utwory Sherwooda Andersona są na tyle krótkie (przeciętny czas czytania pojedynczego opowiadania to jedna godzina), aby nie stanowiło to bariery dla respondentów, którzy odpowiadając na pytania badacza, musieli przeczytać tekst.

Konkretne metody komputerowe oraz sposób zbierania informacji o perspektywach czytelniczych zostały stworzone tak, aby można było powtórzyć badanie na każdym tekście, jednakże ze względu na powyższe przesłanki zostały one przeprowadzone właśnie na opowiadaniach o Winesburg.

6 DK. Elson, N. Dames, KR. McKeown, *Extracting Social Networks From Literary Fiction*, w: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics 2010.

7 WL. Phillips, *How Sherwood Anderson Wrote Winesburg, Ohio*, *American Literature*, Vol. 23, No. 1, March 1951.

3) Postrzegania działania społecznego

3a) Operacjonalizacja badania

W trakcie ustalania metod badawczych rozważaliśmy różne sposoby rozumienia interakcji przed ustaleniem ostatecznej wersji wysłanej do respondentów. Wynikiem był kompromis między chęcią zebrania jak największej ilości informacji, a naukowymi standardami sprawdzalności oraz porównywalności. W związku z eksploracyjnym charakterem badań przedstawiamy kilka kontekstów, które również powinny zostać rozwinięte. Ostatecznie do naszych szczegółowych badań wybraliśmy sieci, tak jak widzi je czytelnik, bez dodatkowych kontekstów. W tym celu zostały przygotowane dwa zadania. Pierwsze, nazywane później komunikacyjnym, polegało na zaznaczeniu interakcji w trakcie czytania tekstu, a drugie – interakcyjne, na wypełnieniu macierzy interakcji wartościami wskazującymi na ważność relacji (znowu rozumianą tak, jak postrzega ją czytelnik). Ze względu na charakter badań nie udało nam się ująć wszystkich aspektów, postanowiliśmy jednak zasygnalizować je w tym artykule.

3b) Podejście komunikacyjne

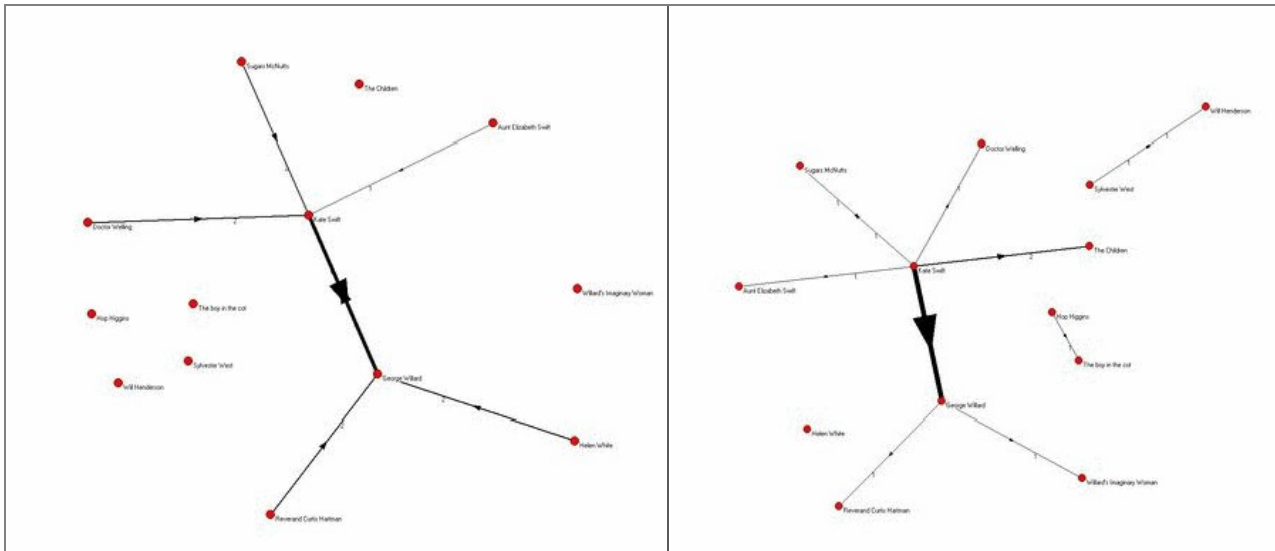
Zaczęliśmy od teorii informacji i schematu:

Nadawca - Komunikat - Odbiorca

Oznacza to istnienie kierunku w kanale komunikacji. Czasami wydaje się trudne do ustalenia dokładnie, kto jest kim w tej konstrukcji, ale przy pewnym rygorze jest to możliwe (mniej lub bardziej obiektywnie). Należy mieć na uwadze, że komunikat nie jest tylko wypowiedzią ustną bądź pisemną. Rozróżniamy również komunikaty niewerbalne, a także różne fizyczne i nie-fizyczne formy interakcji. Założyliśmy, że nadawca to osoba, która ma wpływ na odbiorcę. Przykładowe zdarzenie: jeśli myślę teraz o mojej dziewczynie i jej słowach z poprzedniego dnia, to ma ona wpływ na mnie w tym momencie. Należy rozumieć „myślenie” również jako komunikację, w której odbieram informacje, uprzednio wysłane przez moją dziewczynę. Jako konsekwencję możemy wyróżnić bezpośrednio (explicite) i pośrednio (implicite) interakcje. Tak więc ja, myśląc o mojej dziewczynie, będę odbierał wiadomość implicite w przeciwieństwie do sytuacji, kiedy ona mówiła do mnie dnia poprzedniego, wysyłając komunikat explicite.

Przykładowa sieć ukazująca interakcje, rozumiane w kontekście komunikatów [Ryc.1], została wygenerowana przez jedną osobę, która notowała interakcje w trakcie czytania opowiadania „The Teacher”. Grubość połączenia jest sumą wszystkich zauważanych interakcji, a wielkość strzałki wskazuje na ilość interakcji w danym kierunku. Utwór opowiada o romansie tytułowej nauczycielki z miejscowym dziennikarzem, jest to widoczne poprzez grubość połączenia między nimi. Natomiast, poprzez kierunek interakcji można zauważyć, że to nauczycielka jest aktywną stroną w związku. W tym kontekście, pojedyncza interakcja explicite zaistniałaby, gdyby np. dwoje ludzi rozmawiało ze sobą. Ukryte działanie zostałoby oznaczone jako implicite, np. gdy jedna z postaci

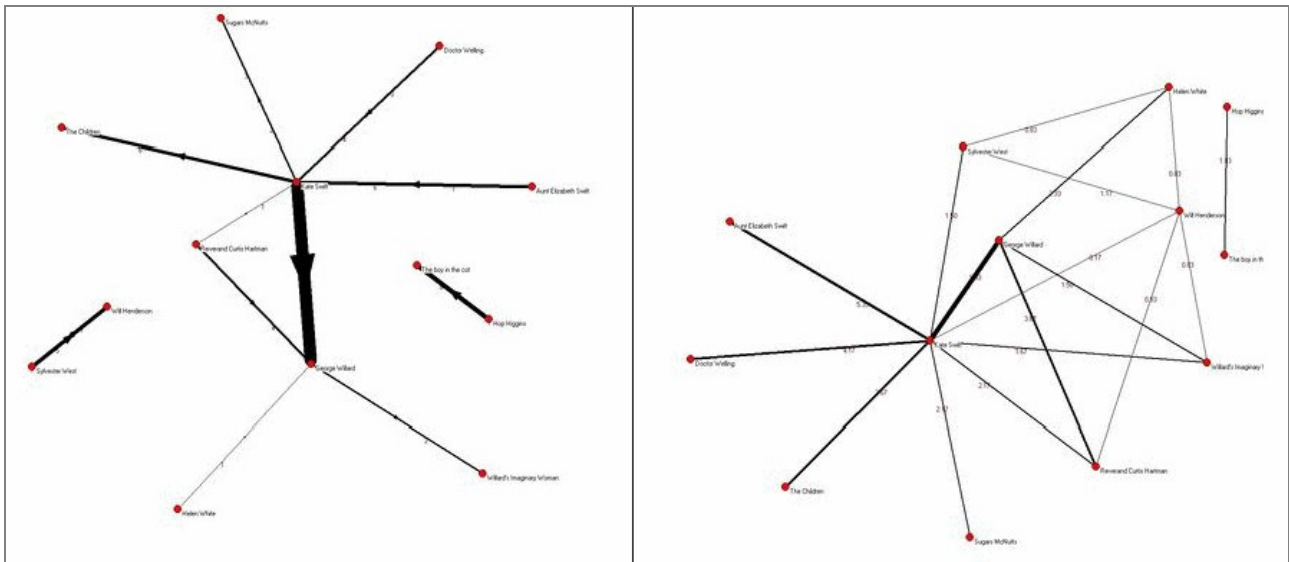
coś sobie przypomiała. Rozróżnienie między zdarzeniami explicite i implicite zostało one usunięte z pełnego badania pilotażowego, pomimo że pozwalałoby na postawienie hipotezy do weryfikacji, że interakcje, które możemy „zobaczyć” są oceniane jako ważniejsze. Przeprowadzenie pogłębionych badań w tym zakresie wymagałoby przyjęcia innej metodologii, mocniej akcentującej czynnik humanistyczny. Pozostawiliśmy więc tą sprawę otwartą.



Ryc. 1) Interakcje implicite (po lewej) vs explicite (po prawej).

3c) Podejście komunikacyjno-interakcyjne

W celu uzupełnienia informacji przygotowano pytanie zadane po przeczytaniu tekstu, w którym czytelnik miał zaznaczyć wagę połączeń między postaciami. Zestawione sieci [Ryc.2] dla zadania komunikacyjnego i interakcyjnego wyraźnie się różnią (są to zebrane wyniki badania pilotażowego). W ostatecznej wersji zrezygnowaliśmy z sieci skierowanej a komunikat został zrzucony na interakcję. Było to spowodowane brakiem możliwości odczytania kierunku przesyłania komunikatu przez utworzone przez nas algorytmy.



Ryc. 2) Komunikaty (po lewej) oraz interakcje (z prawej) z badania pilotażowego.

4) Komputerowe metody ekstrakcji sieci

W implementacji algorytmów staraliśmy się skupić na jak najbardziej „ludzki” sposobie wykrywania interakcji. Jak wyjaśniono wcześniej, interakcja jest to coś więcej niż rozmowa, może być fizyczna, psychiczna itp. Pojawienie się dwóch bohaterów w określonej jednostce tekstu (akapit, zdanie) wydało nam się prostym i dobrym wskaźnikiem szeroko zdefiniowanej interakcji między postaciami. Znacznie trudniejszym zadaniem było wprowadzenie istotności czyli wagi połączenia. Jest to nowy element nie pojawiający się we wcześniejszych publikacjach. To co dla człowieka wydaje się banalne, nie daje się łatwo zaimplementować. Ostatecznie wagę interakcji uzyskuje się przez pomnożenie częstotliwości nazw postaci występujących w danym fragmencie (jednostce analizy). Kolejnym uproszczeniem było określenie kto jest postacią a kto nie. Przykładowo: czy wyimaginowaną przyjaciółkę dziennikarza uznać za bohaterkę? W końcu sami przygotowaliśmy arbitralnie listę bohaterów z aliasami. Należy pamiętać, że algorytm może pracować w dwóch trybach:

- zdaniowym, kiedy jednostkami analizy są poszczególne zdania;
- akapitowym, kiedy jednostkami analizy są poszczególne akapity.

5) Specyfikacja badania

Badanie zostało przeprowadzone w postaci ogólnodostępnej ankiety. Respondenci zostali poproszeni o wypełnienie czterech stron: formularza zgody, pierwszego zadania [Ryc. 3], drugiego zadania [Ryc. 4] i metryczki. W zadaniu pierwszym i drugim istotność interakcji była oceniana w zakresie [0-10]. W metryczce pytaliśmy się o różne czynniki demograficzno-społeczne jak: wiek, płeć, czy rodzaj wykształcenia. Do głównego badania wybraliśmy opowiadanie „The Philosopher”,

w którym opisywane są perypetie młodego lokalnego dziennikarza. Osia fabuły są tu rozmowy wspomnianego głównego bohatera z jego przyjacielem, lekarzem, na temat ich doświadczeń życiowych. Analiza wyników była anonimowa, ale badani mieli szansę zostawić nam adres kontaktowy. Ostatecznie ankietę w całości wypełniło 36 osób (niestety mieliśmy bardzo duży odsetek odrzuceń, gdyż ponad 200 razy ankieta była rozpoczynana), głównie z Wielkiej Brytanii i Stanów Zjednoczonych. W większości byli to studenci, wśród których sporą część stanowili nienatywni użytkownicy języka angielskiego (w tym języku badanie zostało przeprowadzone), chociaż znali oni ten język przynajmniej w stopniu umożliwiającym swobodne czytanie.

[Previous entries](#) [\[want a blank form?\]](#)

character 1	character 2	importance	delete
George Willard	Kate Swift	4	[del]
Kate Swift	George Willard	3	[del]
Kate Swift	George Willard	7	[del]
Kate Swift	The Children	1	[del]

As you read mark down any

You will be asked to distinguish between interactions that are 'important' and 'unimportant' according to your reading of the story.

choose characters: importance

Ryc. 3) Formularz do wypełniania zadania pierwszego komunikacyjnego z najważniejszymi wytycznymi

At the end of the story, use the table to rate out of ten the overall importance of the interactions between each set of two characters based on your reading of the story.

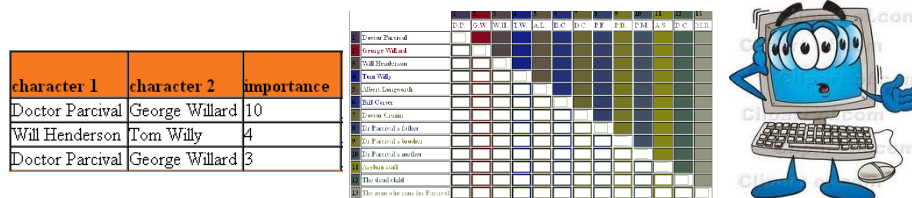
		1	2	3	4	5	6	7	8	9	10	11	12	13
		W.H.	S.H.	G.W.	K.S.	W.I.	H.W.	H.H.	B.C.	C.H.	D.W.	T.C.	S.M.	E.S.
1	Will Henderson	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2	Sylvester West	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3	George Willard	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4	Kate Swift	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5	Willard's Imaginary Woman	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
6	Helen White	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
7	Hop Higgins	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
8	The boy in the cot	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
9	Reverend Curtis Hartman	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
10	Doctor Welling	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
11	The Children	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
12	Sugars McNutts	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
13	Aunt Elizabeth Swift	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Ryc. 4) Formularz do wypełniania zadania drugiego interakcyjnego z najważniejszymi wytycznymi

6) Analiza danych

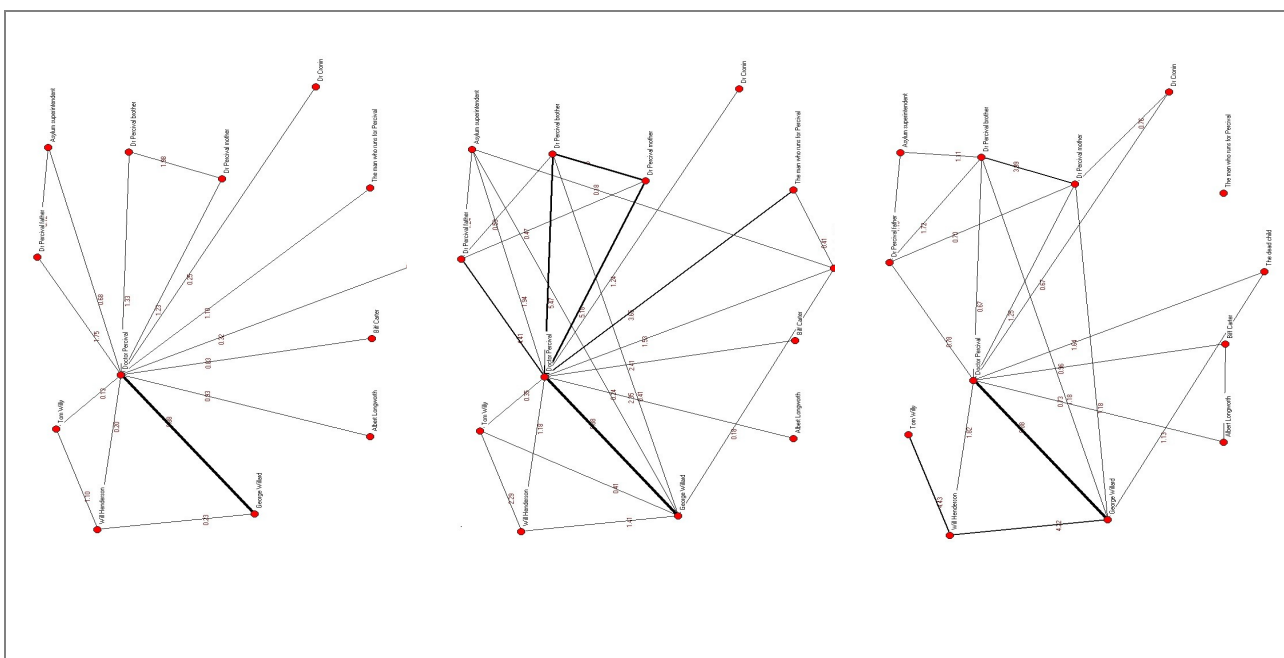
6a) Budowa sieci

W celu porównania powyższych metod (dwóch ludzkich: pierwszej komunikacyjnej i drugiej interakcyjnej oraz dwóch komputerowych: akapitowa i zdaniowa) dokonano normalizacji wyników [Ryc. 5].



Ryc. 5) Wizualizacja problemu różnego rodzaju wyników otrzymanych różnymi metodami.

Uzyskane sieci można przedstawić graficznie [Ryc. 6]. Widać, że w ogólności wyglądają one podobnie. Najważniejsze połączenie z punktu widzenia fabuły (pomiędzy dziennikarzem a doktorem) w każdej metodzie jest uwypuklone. Zauważalną różnicą jest za to różna ilość połączeń w sieciach. Zadanie pierwsze-komunikacyjne wykazuje najmniej połączeń (16) a drugie-interakcyjne najwięcej (23), podczas gdy algorytm akapitowy daje wynik pośredni (21).



Ryc. 6) Uzyskane sieci społeczne opowiadania „The Philosopher” z zadania pierwszego-komunikacyjnego (po lewej), z zadania drugiego-interakcyjnego (po prawej) oraz algorytmu akapitowego.

6b) Korelacja między metodami

Najważniejszym punktem projektu było porównanie wyników uzyskanych od respondentów oraz za

pomocą algorytmów [Tab. 1]. Zauważmy, że algorytm zdaniowy lepiej koreluje z zadaniem komunikacyjnym, a algorytm akapitowy lepiej z zadaniem interakcyjnym. Mimo to sumarycznie nieznacznie lepiej wypada algorytm akapitowy w porównaniu ze zdaniowym.

	Algorytm akapitowy	Algorytm zdaniowy
AD 1 (komunikacyjna)	0,84	0,91
AD 2 (interakcyjna)	0,70	0,58

Tab. 1) Macierz korelacji pomiędzy sieciami otrzymanymi od respondentów i za pomocą algorytmu.

6c) Korelacje pomiędzy odpowiedziami respondentów

Porównajmy teraz korelacje pomiędzy sieciami uzyskiwanymi przez tych samych respondentów w zadaniu pierwszym-komunikacyjnym i drugim-interakcyjnym. Zgodność odpowiedzi spróbowaliśmy wyjaśnić za pomocą zmiennych niezależnych podanych przez respondenta w metryczce za pomocą regresji logistycznej zmiennych o wymiarze nominalnym [Tab.2]. Okazało się, że tylko rodzaj wykształcenia ma istotny statystycznie wpływ na te korelacje, a dokładnie tylko respondenci o wykształceniu ścisłym wyraźnie odpowiadali w podobny sposób. I tak osoba z wykształceniem ścisłym wykazuje się większą powtarzalnością pomiędzy oboma zadaniami.

wykształcenie	parametr	p-Value
Arts&humanities	0,05	0,37
Social science	-0,03	0,71
Science	0,16	<0,01

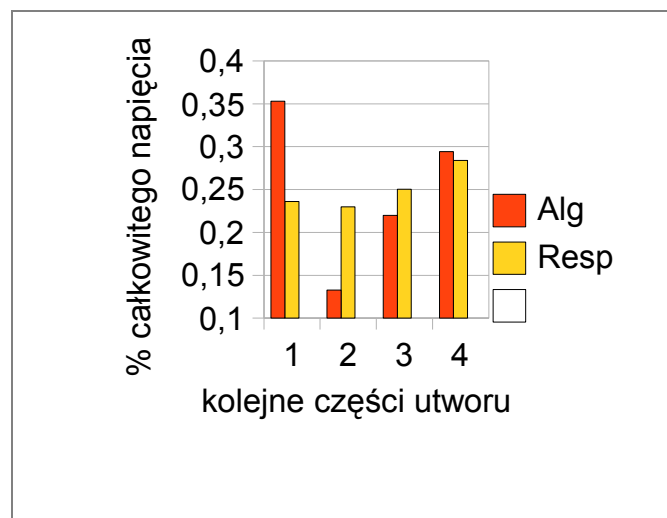
Tab. 2) Regresja logistyczna zmiennych nominalnych pokazująca relację między wykształceniem ścisłym a zgodnością w uzupełnianiu zadania pierwszego-komunikacyjnego oraz drugiego-interakcyjnego.

6d) Testowanie tworzenia trójkąta Freytag'a (punkt kulminacyjny)

Ostatnim punktem naszej analizy było zbadanie zdolności dynamicznych własności sieci do odzwierciedlenia kulminacji napięcia charakterystycznego dla większości utworów literackich.

Polegało to na zbadaniu intensywności i znaczenia interakcji w czasie. Nasza hipoteza zakładała, że suma ważonych interakcji będzie odpowiadać budowaniu napięcia w narracji. W tym celu sprawdziliśmy czy charakterystyczny trójkąt Freytag'a może być otrzymany automatycznie. Dzięki

temu mielibyśmy metodę ilościową, a wynik jakościowy. W tym celu wykorzystaliśmy dane z zadania pierwszego-komunikacyjnego oraz wyniki algorytmu akapitowego. Podzieliliśmy zebrane dane na cztery partie zgodnie z chronologią czasową z tym, że dla metody komputerowej ten podział miał miejsce wedle ilości wierszy, a z danych ankietowym wydzieliśmy równoliczne kwartyle. Po utworzeniu wykresu z danych [Ryc.7], okazało się, że nie otrzymaliśmy oczekiwanej formacji. Możemy interpretować to na różne sposoby. Np. komputer znajduje wiele interakcji między postaciami na początku, ponieważ autor przedstawia większość bohaterów do akcji właśnie w tym miejscu. Respondenci za to w miarę jednorodnie oznaczają intensywność interakcji, z tym że obserwujemy niewielką tendencję wzrostową wraz z upływem czasu.



Ryc. 7) Konstrukcja anty-trójkąta przy pomocy interakcji zebranych od respondentów w zadaniu pierwszym-komunikacyjnym oraz wygenerowanych przez algorytm akapitowy.

7) Wnioski i spekulacje

Odkrywanie sieci społecznych w literaturze okazało się pasjonującym tematem, w którym wiele jest jeszcze do zrobienia, ponieważ nadal jest to pionierski obszar badań. Opracowaliśmy metody pozyskiwania sieci na podstawie ankiet czytelników, które mogą zostać wykorzystane do analizy innych dzieł niż „Winesburg, Ohio”, choć już na podstawie uzyskanych wyników możemy wyciągnąć pewne wnioski i postawić nowe hipotezy badawcze. W naszych badania uprościliśmy sposób zbierania informacji od czytelnika w celu zachowania możliwości porównania z metodami komputerowym, m.in. dlatego nie umieściliśmy respondenta w żadnym konkretnym kontekście czytelniczym. Już przedwstępna analiza pokazała, że sposób zadania pytań w ankiecie ma znaczenie. Co więcej, zamiast nadawać wagę interakcji tak jak ją widzi czytelnik, można by narzucić orientację choćby na fabułę. W ten sposób percepcja czytelnika mogłaby być ukierunkowana na świat przedstawiony w utworze a nie na jego subiektywną wizję tegoż świata. Być może zmieniłoby to uzyskane wyniki i zbliżyłoby to metody ankietowe z komputerowymi.

Pozostaje to nadal otwartym zagadnieniem, podobnie jak pytanie o związek między interakcjami typu *implicite* i *explicite*. Jednak odnosząc się do uzyskanych przez nas wyników najważniejsze wydają się korelacje między rezultatami ankiet oraz algorytmów [Tab. 1]. Niezmiernie ciekawym aspektem jest inny poziom korelacji odpowiednich metod komputerowych z różnymi zadaniami czytelniczymi. Wskazuje to na niejednoznaczność w sposobach opisywania interakcji i czułość na czynniki zewnętrzne. W skrócie oznaczać to może, że nie istnieje jedna uniwersalna metoda ukazująca sieć społeczną w utworze literackim. Interesującym wynikiem było też pokazanie związku między sposobem postrzegania sieci a wykształceniem (reprezentanci nauk ścisłych budują sieci bardziej powtarzalne niż reszta). Na koniec sprawdziliśmy, że mierząc intensywność interakcji nie możemy przenieść tego bezpośrednio na budowanie napięcia w utworze. Podsumowując wydaje nam się, że poruszyliśmy wiele wątków jakie niesie ze sobą wykorzystanie analizy sieci społecznych w literaturze, która ze względu na charakter ilościowy może być narzędziem uzupełniającym w warsztacie współczesnego humanisty.

8) Podziękowania

Projekt został zrealizowany w ramach grantu Transit Interdisciplinary Project na Uniwersytecie w Yorku (Wielka Brytania). Nie udało się przeprowadzić badań bez pomocy merytorycznej wielu naukowców z Yorku a przede wszystkim: Richarda Welsha (filologia angielska), Elvy Robinson (biologia), Dana Franka (informatyka) oraz Johna Forrestera (antropologia).

Streszczenie

W naszym interdyscyplinarnym projekcie pragniemy wykorzystać metody analizy sieciowej, aby lepiej zrozumieć sposób kreacji oraz przedstawienia świata przez autorów utworów literackich. Jednakże percepcja takiego świata zależy od subiektywnej wizji czytelnika, więc położyliśmy szczególną uwagę na różne sposoby ekstrakcji sieci powiązań społecznych z fikcyjnej rzeczywistości. Celem naszych badań było odczytanie różnych interakcji społecznych w tekście przez porównanie sieci otrzymanych przez algorytmy przetwarzania języka naturalnego (NLP) z tymi odtworzonymi na podstawie kwestionariuszy wypełnionych przez czytelników. Sieci dialogów czy uczestnictwa w tym samym fragmencie tekstu zostały już opisane przez amerykańskich naukowców, ale wciąż brakowało analizy relacji na ogólniejszym poziomie. Zaproponowaliśmy kilka metod NLP w celu detekcji tych interakcji i skonfrontowaliśmy je z ludzkim postrzeganiem. Przy okazji odkryliśmy obszary teorii literatury, w których nie da się wykorzystać analizy sieciowej (np. interakcje nawiązujące do fabuły nie tworzą klasycznego trójkąta z punktem kulminacyjnym).

Streszczenie po angielsku

We want to make a cross-disciplinary leap, and use the tools of network theory to understand and explore narrative structure in literary fiction, a still under-utilized approach. However, the systems in fiction are sensitive on reader's subjectivity and attention must to be paid to different methods of extracting networks. The project aims to investigate the different ways social interactions are 'read' in texts by comparing networks produced by automated algorithms-natural language processing (NLP) with those created by surveying more subjective human responses. Conversation networks from fiction have been already extracted by scientists, but the more general framework surrounding these interactions was missing. We propose several NLP methods for detecting interactions, and test them against a range of human perceptions. In doing so, we uncovered some limitations of using network analysis to test literary theory (e.g. interaction, which correspond to the plot, do not form climax).

Noty biograficzne

Andrzej Jarynowski jest doktorantem fizyki na Uniwersytecie Jagiellońskim. Podzodzi z Pomorza. Studiował m.in. matematykę na Politechnice Wrocławskiej i socjologię na Uniwersytecie Wrocławskim. Zajmuje się wykorzystaniem teorii układów złożonych w naukach społecznych. Obecnie poszukuje zastosowań analizy sieci społecznych w literaturze i w muzyce.

Stephanie Boland ukończyła filologię angielską na Uniwersytecie w Exeterze, a obecnie prowadzi badania nad współczesną literaturą na Uniwersytecie w Londynie (Queen Mary). Jej głównym obszarem zainteresowań są związki regionalne w opisie społeczności lokalnych m.in. u Jamesa Joyce'a. Podobnie jak jej ulubiony pisarz jest z pochodzenia Irlandką.