

Każdego dnia miliardy wiadomości, komentarzy, artykułów, postów na blogach, e-maili, tweetów i innych rodzajów informacji jest wymieniane we różnych rodzajach sieci społecznościowych. Badania nad rozpowszechnianiem informacji stały się bardzo owocne i w szczególności znajdują zastosowania do maksymalizacji wpływu i wiralności plotek lub ulepszenia algorytmów routingu. Warto zwrócić również uwagę na to, że modele rozpowszechniania informacji oparte są na klasycznych modelach epidemiologicznych, np. SIR. W ostatnich miesiącach modele te były szeroko stosowane do modelowania rozprzestrzeniania się wirusa COVID.

Dlatego zrozumienie i modelowanie procesów wirusowych stało się kluczowym kierunkiem badań. Modele opisujące procesy wirusowe zwykle zakładają, że proces ten jest stochastyczny, np. słynny model SIR. Wydaje się, że model ten jest poprawny dla przypadku, dla którego został stworzony, tj. do opisania procesu rozprzestrzeniania się chorób. Jednak, jak pokazaliśmy w naszej poprzedniej pracy (HT 2016) - nie dotyczy on rozpowszechniania informacji. Po pierwsze, model ten nie uwzględnia faktu, że informacje stają się z czasem mniej aktualne, a ludzie udostępniają je mniej aktywnie. Po drugie, informacje są przekazywane różnymi kanałami. Jeśli chcemy badać rozpowszechnianie się informacji w sieci Twitter, musimy również wziąć pod uwagę inne media, np. środki masowego przekazu. W szczególności brak tych efektów powoduje, że model SIR przecenia prawdopodobieństwo, że informacja stanie się wiralna, tj. dotrze do prawie całej sieci. Nasza praca (HT 2016) wyjaśnia obserwowane rozmiary kaskady, uwzględniając te dwa efekty:

- wykładniczy spadek prawdopodobieństwa dalszego rozprzestrzeniania się pogłoski,
- wieloźródłowy charakter procesu odpowiadający za rozprzestrzenianie się plotek poza siecią Twitter.

Inne możliwe wyjaśnienie można znaleźć w naszej pracy (WWW 2017), w której stworzyliśmy pierwszy model teoretyczny, który wyjaśnia, dlaczego rozkład wielkości kaskady jest zgodny z prawem potęgowym. W naszej pracy wprowadziliśmy pojęcie kierunku rozprzestrzeniania się informacji, tj. z węzłów wysokiego stopnia i wysokiego zaufania. Motywacją tego założenia jest fakt, że ludzie częściej dzielą się informacjami pochodzącymi z węzłów, które ma wielu znajomych. Innymi słowy, wydaje się, że tak naprawdę nie jesteśmy jeszcze daleko od dobrego zrozumienia mechanizmu rozpowszechniania się informacji w sieciach społecznościowych. Dzieje się tak pomimo faktu, że w sieciach społecznościowych procesy kaskadowe można bardzo dokładnie prześledzić. Nasz brak zrozumienia oznacza, że nie jesteśmy w stanie poprawnie oszacować ryzyka związanego z bardzo rzadkimi zdarzeniami. W szczególności, o ile nam wiadomo, nasz artykuł (HT 2016) jest jedynym przypadkiem, w którym zastosowano miarę, która prawidłowo uwzględnia rzadkie zdarzenia. Rodzi to pytanie, czy w zastosowaniach epidemiologicznych takich modeli rzadkie zdarzenia, np. w rozprzestrzenianie się pandemii COVID, są poprawnie opisane.

Kolejną linią badań procesów wiralowych jest przewidywanie popularności danej informacji. Należy zauważyć, że modele te tworzone są w oparciu o zupełnie inne podejście niż zakładane w naszych pracach. Typowym podejściem jest zbudowanie modelu regresji, który w oparciu o obserwowane cechy procesu przewiduje jego dalszą ewolucję. Modele te mają jednak ograniczoną skuteczność, ponieważ pośrednio zakładają, że proces jest deterministyczny, a przecież ma on charakter stochastyczny i jego ewolucja nie jest z góry określona. Dlatego wyzwaniem jest opracowanie modeli, które przewidywałyby wszystkie możliwe kontynuacje ewolucji opisane jako dystrybucja. W szczególności tylko takie podejście może prowadzić do statystycznie poprawnych wyników, które przewidywałyby szanse, że proces dotrze do całej sieci. To prowadzi do wyzwań badawczych, które tworzą tematy dla zadań niniejszego wniosku:

- naszym celem jest wskazanie mechanizmów odpowiedzialnych za zanik prawdopodobieństwa dalszego udostępniania informacji.
- będziemy stochastycznie modelować ryzyko, że proces wirusowy rozprzestrzenia się na całą sieć lub na jej znaczną część.
- będziemy pracować nad modelami do przewidywania ewolucji konkretnej plotki.
- jak wywnioskować parametry i sposoby transmisji procesu wirusowego z pośrednich obserwacji.
- sprawdzimy, czy możliwe jest wykrycie natury wirala, np. czy związany on jest ze zdarzeniem w świecie rzeczywistym, fałszywą wiadomością lub artykułem naukowym.
- zastosujemy metodologię opracowaną w tym projekcie do modelowania epidemii COVID. W szczególności nasza praca rzuci światło na prawidłowe opisanie ryzyka tego procesu.