

O czym piszą piraci?

- analiza wypowiedzi w serwisie TorrentFreak

Alina Niemiro-Hardy i Wojciech Hardy

Wstęp

Wielu badaczy podjęło się analizy tzw. 'piractwa' internetowego, przyglądając mu się tak od strony jego wpływu na rynek, jak i analizując profil i motywacje osób podejmujących się wymiany plikami. Jak można się spodziewać, najlepszym źródłem informacji o tzw. piratach jest sam Internet, w którym przeprowadzają oni większą część swojej działalności. Wiele badań mierzy ruch w Internecie, oraz przygląda się przepływowi w tzw. sieciach Peer-To-Peer.

Interesującym jest jednak co ludzie sami mają do powiedzenia o piractwie. Choć wielu naukowców podejmowało się przeprowadzenia rozbudowanych ankiet na ten temat, jest to obarczone pewnym ryzykiem. Przede wszystkim, osoby biorące udział w badaniu nie muszą zachowywać się w pełni 'naturalnie', co objawić się może chociażby poprzez deklarację poglądów bardziej krytycznych wobec piractwa niż są one w rzeczywistości.

Alternatywnych narzędzi do uchwycenia tego problemu dostarczają metody Text Mining, umożliwiające wyciągnięcie informacji bezpośrednio z naturalnego środowiska, w którym obracają się Internauci, czyli Internetu. Tego właśnie podejmujemy się w tej pracy.

Do analizy problemu wykorzystujemy wypowiedzi użytkowników serwisu TorrentFreak (TF) – portalu dostarczającego regularnych informacji o własności intelektualnej w sieci. Mimo, że charakter wpisów na portalu pozbawiony jest fałszywych przekazów i nadinterpretacji prezentowanych treści, serwis przyciąga głównie środowiska pirackie.

Z pomocą komentarzy zebranych z serwisu TF, zbadaliśmy ogólne zdania dotyczące piractwa najczęściej kopiowanych dóbr cyfrowych. Postanowiliśmy sprawdzić czy wypowiedzi użytkowników serwisu różnią się od siebie w zależności od omawianego rodzaju dóbr - dlatego analiza dokonana została w podziale na trzy grupy: o filmach, muzyce i grach.

1. O danych.

W pracy tej wykorzystano dane zabrane z popularnego serwisu dotyczącego praw własności intelektualnych w sieci. Portal TF jest portalem informacyjnym, w sposób względnie rzetelnie opisującym wszelkie wydarzenia związane z poruszonym zagadnieniem. W serwisie przytaczany jest szeroki wachlarz tematów, w tym nowości dotyczące branż takich jak filmy, muzyka, czy gry komputerowe.

Zebrane do tej pracy dane, to komentarze zamieszczane pod wpisami dotyczącymi właśnie tych trzech kategorii. Zebrane wpisy o muzyce i filmach sięgają jeszcze kwietnia 2012, w przypadku gier zaś sięgnęliśmy także do roku 2011 z uwagi na mniejszą liczbę komentarzy pod tymi wpisami. Komentarze zamieszczane na stronie TF to zwykle niezbyt długie wypowiedzi, często prezentujące jakieś zdanie na temat wpływu piractwa na rynek, organizacji zbiorowego zarządzania, czy też cyfrowych form ochrony plików (tzw. digital restriction management, w skrócie DRM).

Przykładowe wypowiedzi z TF można zobaczyć na obrazku 1.



Rysunek 1. Przykładowe wypowiedzi z serwisu TorrentFreak.

Źródło: <http://torrentfreak.com/top-10-most-pirated-movies-of-the-week-130603/>

Z wstępnego zbioru wypowiedzi, wybraliśmy tylko te, których liczba znaków przekraczała 25. W ten sposób odrzuciliśmy krótkie jednozdaniowe lub jednosłowne odpowiedzi, z których nie wynikała żadna większa myśl i które najczęściej stanowiły odpowiedź do poprzednich komentarzy – nie pogląd sam w sobie. Ze względu na potoczny charakter wpisów na stronie, część słów przewijających się przez tę pracę została ocenzurowana.

Przed rozpoczęciem analizy wszystkie wypowiedzi zostały dodatkowo ‘oczyszczone’ i przygotowane do dalszej pracy. Po pierwsze nie uwzględniliśmy w analizie jakichkolwiek części mowy nie zawierających informacji (jak zaimków, czy spójników). Po drugie zadaliśmy o to, żeby będące synonimami (lub odnoszące się do tego samego jak np.: skróty) były rozpoznawane jako słowa o tym samym znaczeniu. To szczególnie istotne w środowisku o tak zróżnicowanym nazewnictwie – np. po polsku: pobieranie (ang. *downloading*) można też nazwać ‘piraceniem’ (ang. *pirating*) czy kopiowaniem (ang. *copying*).

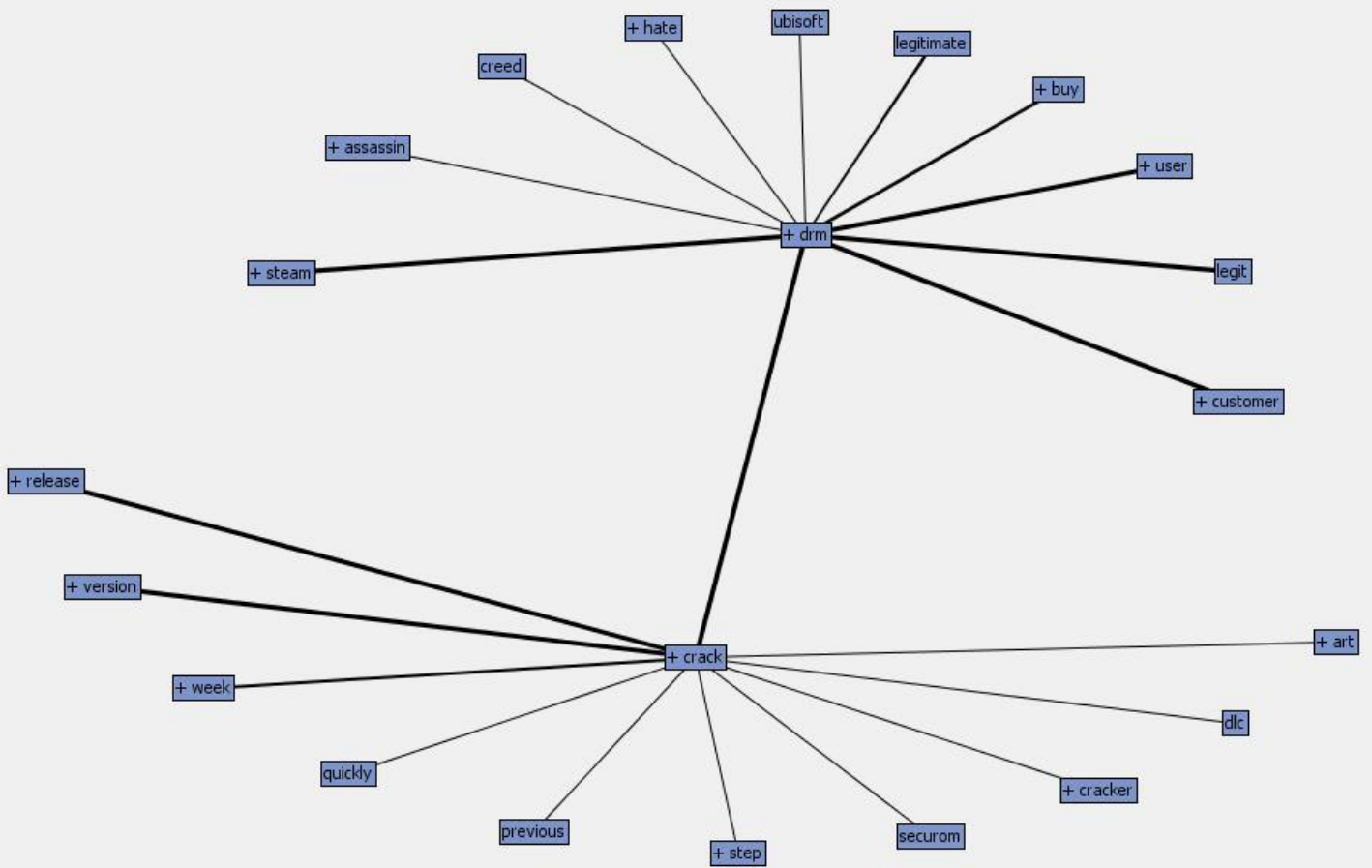
Analizie poddaliśmy ostatecznie trzy zbiory wypowiedzi: o grach (2736 wypowiedzi), o muzyce (3321 wypowiedzi), i o filmach (4813 wypowiedzi). Łącznie zebranych zostało 10870 wypowiedzi.

2. Concept links, czyli powiązania słów

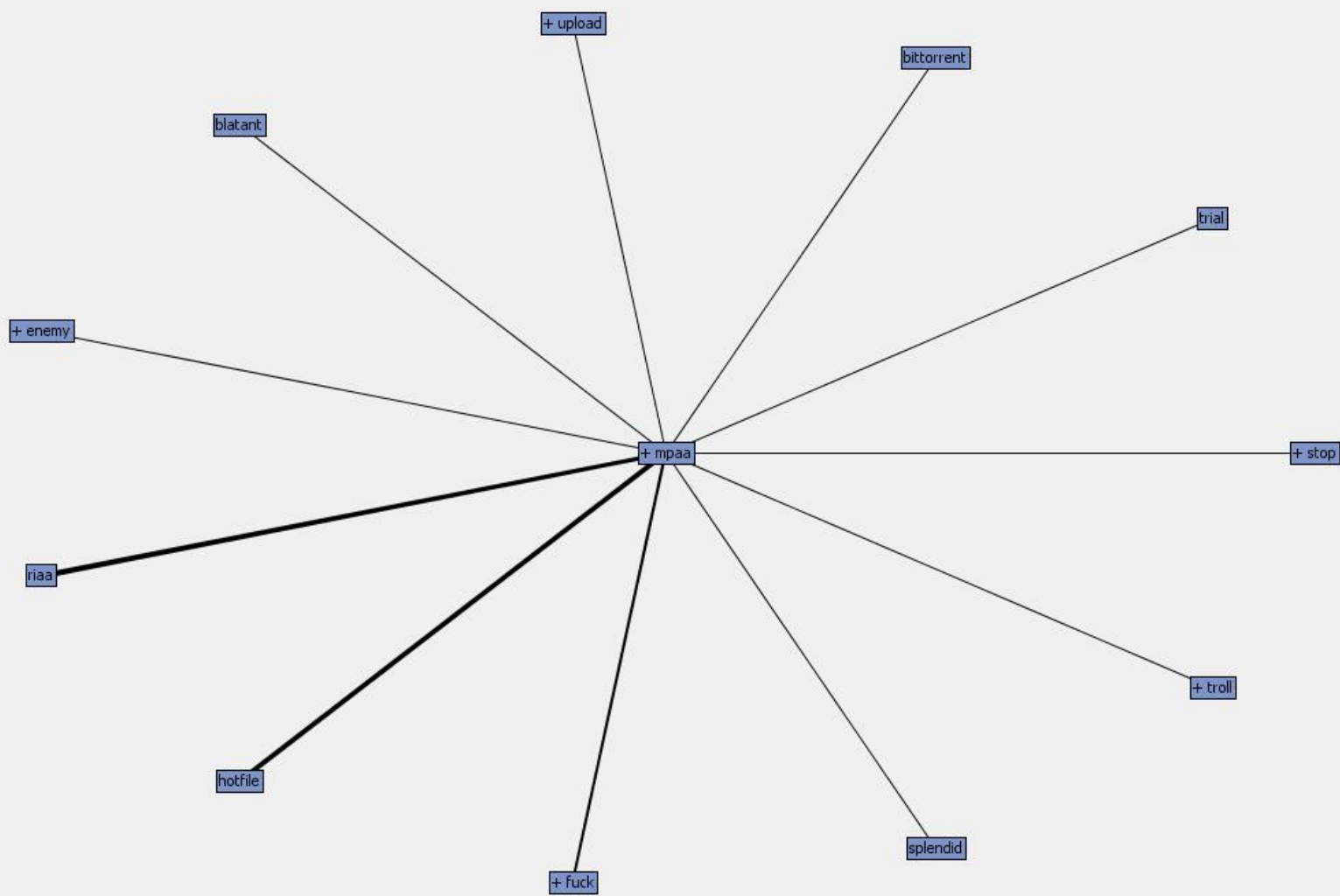
W ramach analizy tzw. *concept linków*, skupiliśmy się na paru najważniejszych zwrotach. Dla każdej gałęzi były to: piractwo-ściągnięcie (*piracy*), piractwo-udostępnianie (*upload*), oraz podmiot związany z ochroną branży, ale często negatywnie postrzegany przez jej konsumentów (organizacje zbiorowego zarządzania MPAA i RIAA dla filmów i muzyki, oraz zabezpieczenia typu DRM dla gier).

2.1. DRM, MPAA, RIAA.

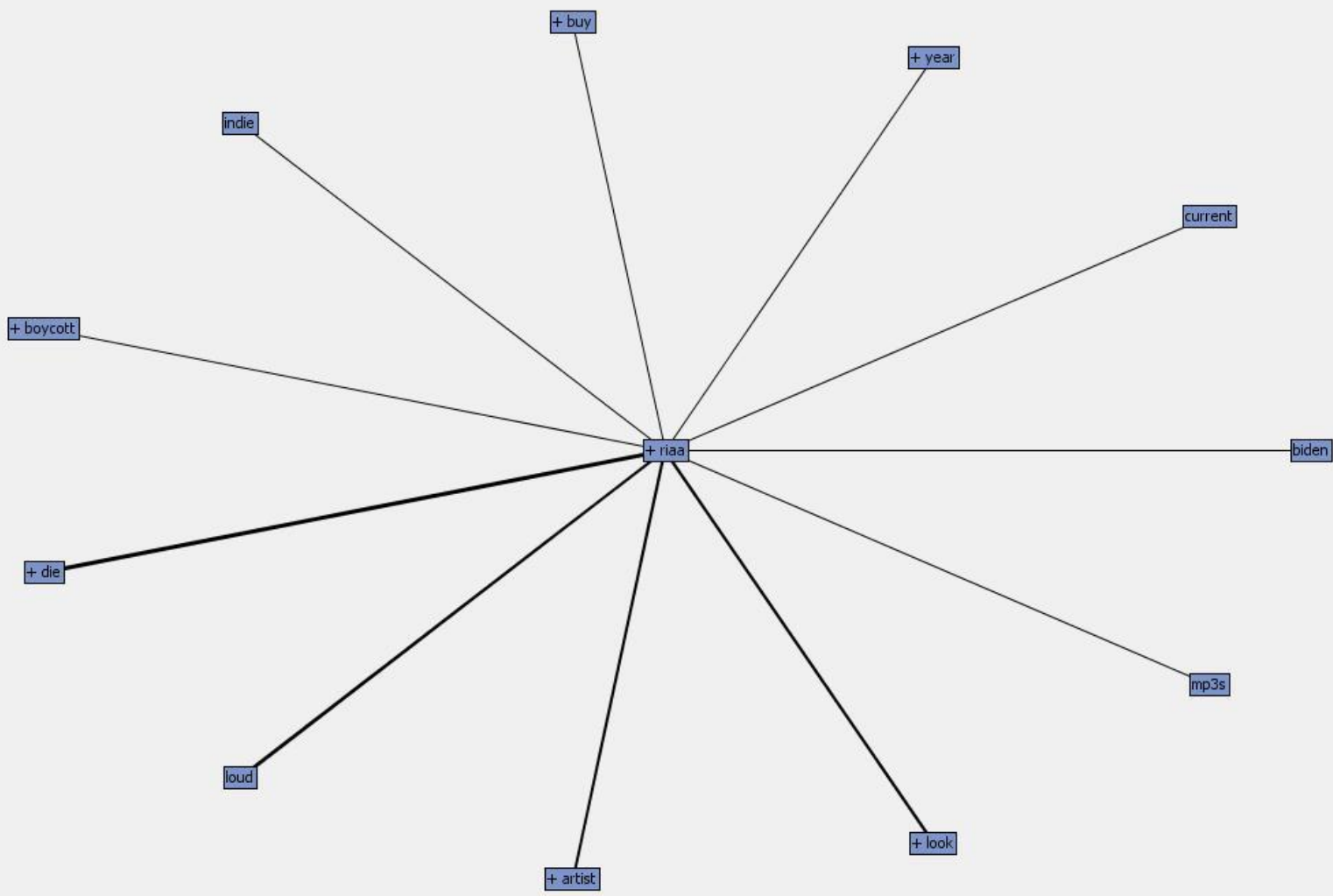
W pierwszym kroku przyjrzymy się wynikom dla MPAA, RIAA i DRM. Jak można zaobserwować na trzech pierwszych wykresach, za równo zabezpieczenia DRM jak i organizacje zbiorowego zarządzania rzeczywiście nie budzą pozytywnych skojarzeń wśród użytkowników serwisu TF. Powiązane z nimi słowa często nacechowane są bardzo negatywnie (np.: *hate, die, enemy, troll, f*ck, czy boycott*). Ciekawą gałęzią jest ta powiązana ze słowem *crack* w przypadku DRM, czyli z procesem łamania cyfrowych zabezpieczeń. Wymienione słowa sugerują powszechność tego czynu, a także często jego szybkość względem premiery gry (*quickly, week, release, cracker, version*).



Rysunek 2. Concept links, tematyka: gry, DRM.



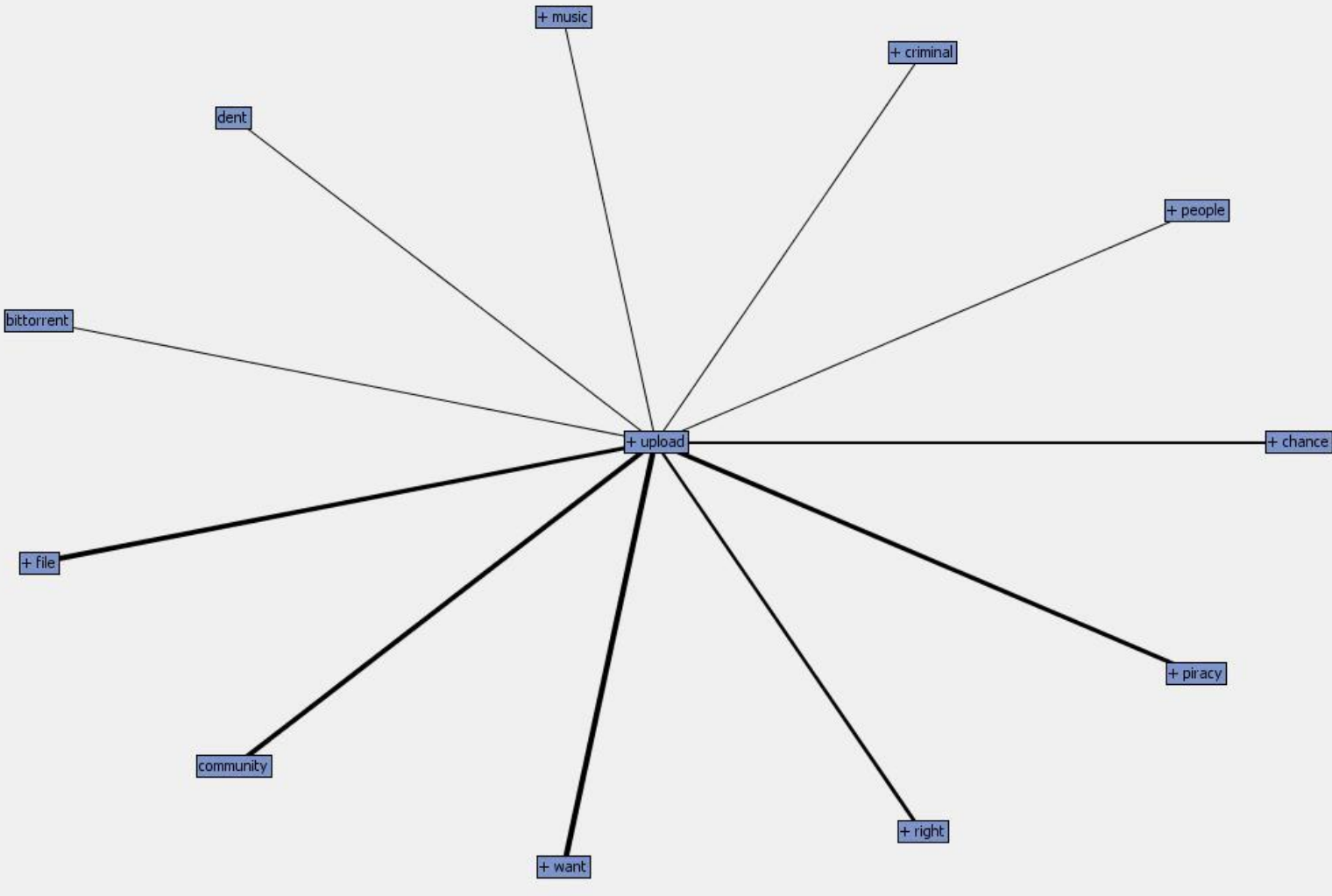
Rysunek 3. Concept links, tematyka: film, MPAA.



Rysunek 4. Concept links, tematyka: muzyka, RIAA.

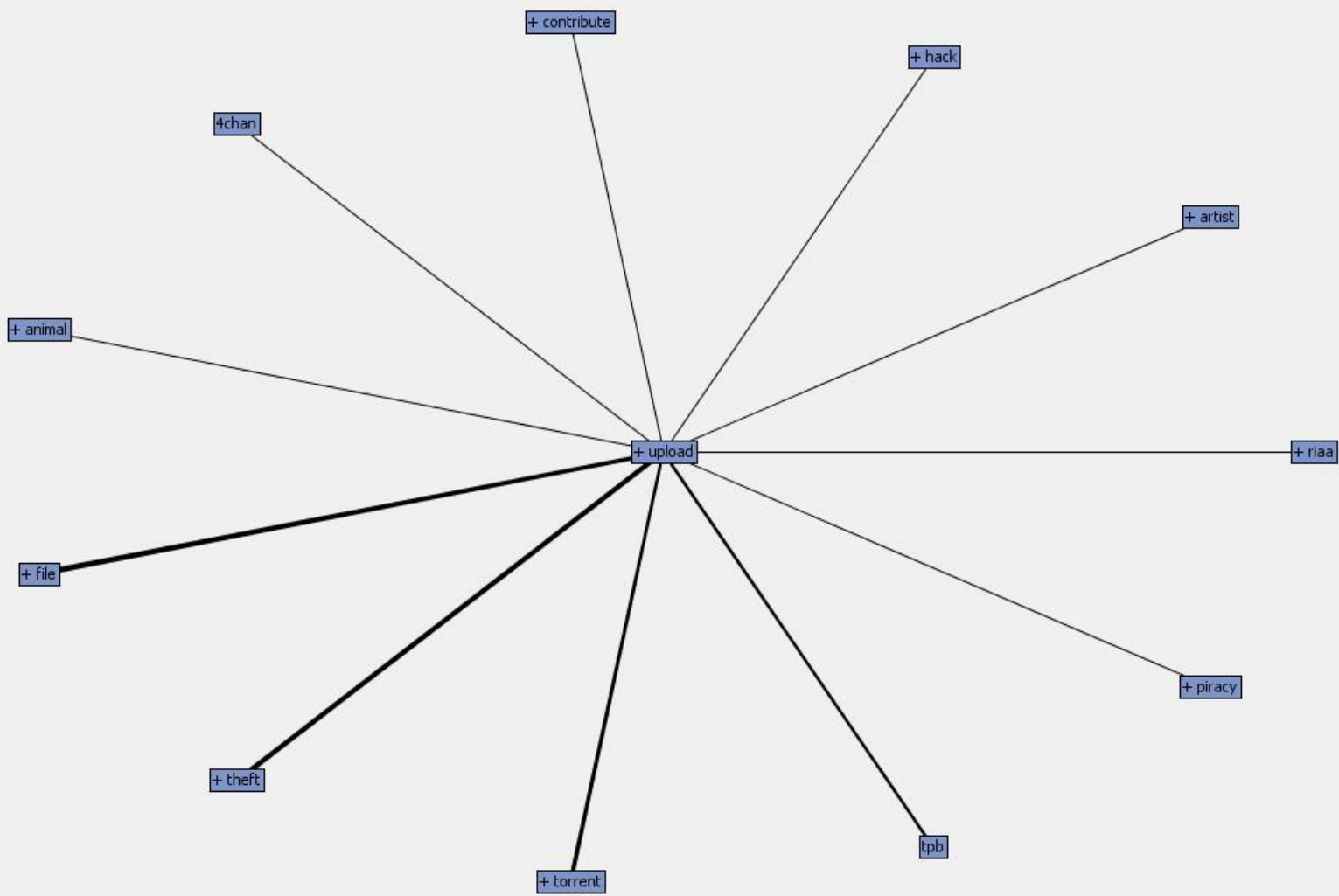
2.2. Upload.

Następnie możemy popatrzeć na słowa często łączone ze słowem *upload*. Motywacje do udostępniania plików mogą znacznie różnić się od tych powiązanych z ich ściąganiem. Dla wszystkich branż, widać powiązanie udostępniania ze społecznością,

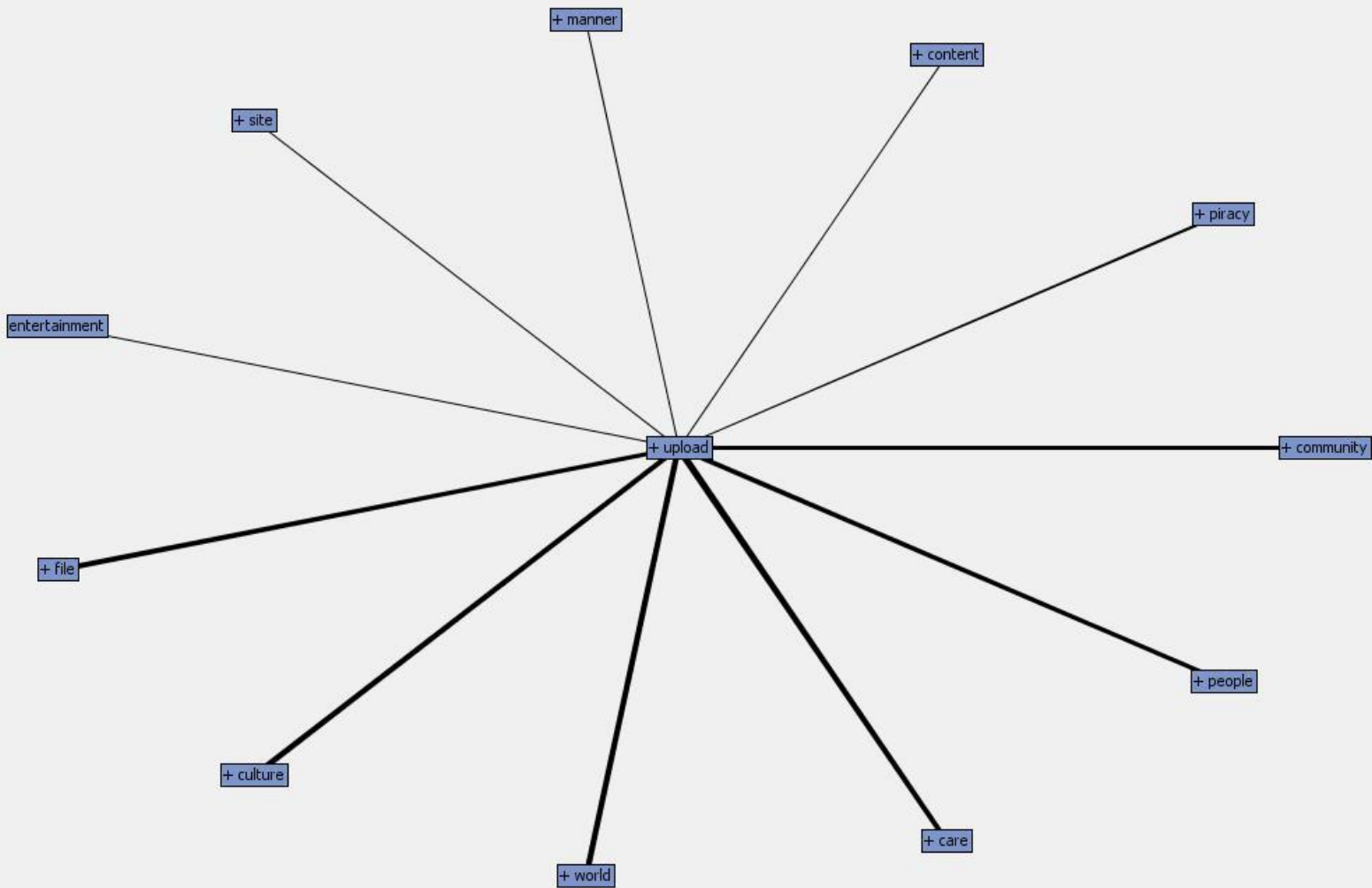


Rysunek 5. Concept links, tematyka: muzyka, upload.

często nawet z ogólną populacją świata. Ta cecha udostępniania jest szczególnie silna w przypadku filmów. W tym wypadku, słowo *upload* łączy się z całą grupą słów wskazujących na niemal charytatywny charakter takiej działalności – szerzenie kultury ze światem. Te słowa to: *community*, *people*, *care*, *world*, *culture*. Nie można oczywiście zignorować także negatywnych powiązań, wyraźnie pojawiających się nawet w środowisku bardziej pro-pirackim. Słowo *upload* często kojarzone jest też z *criminal*, czy *theft*. Te negatywnie nacechowane określenia występują głównie w przypadku branży muzycznej i gier komputerowych.



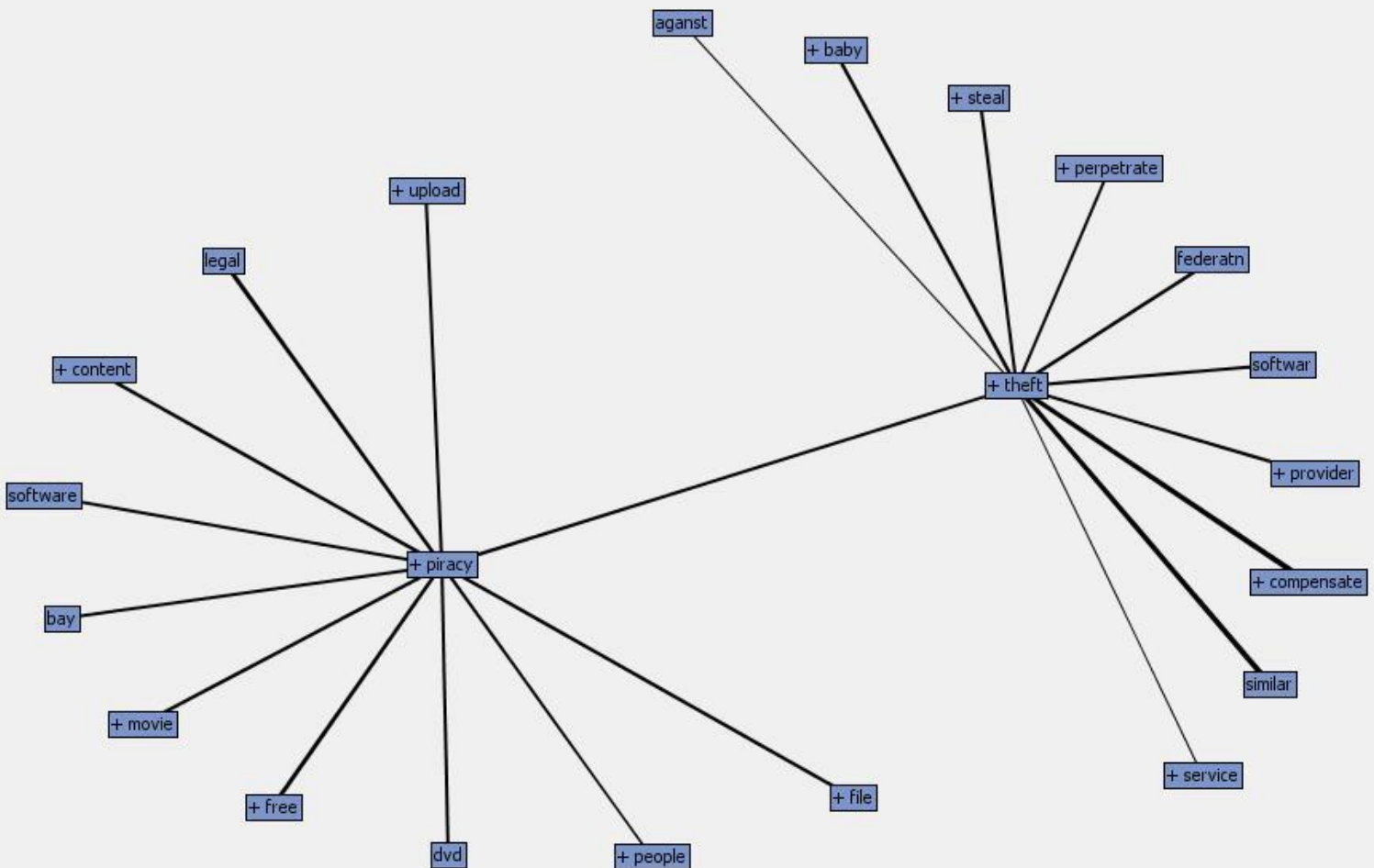
Rysunek 7. Concept links, tematyka: gry, upload



Rysunek 6. Concept links, tematyka: film, upload.

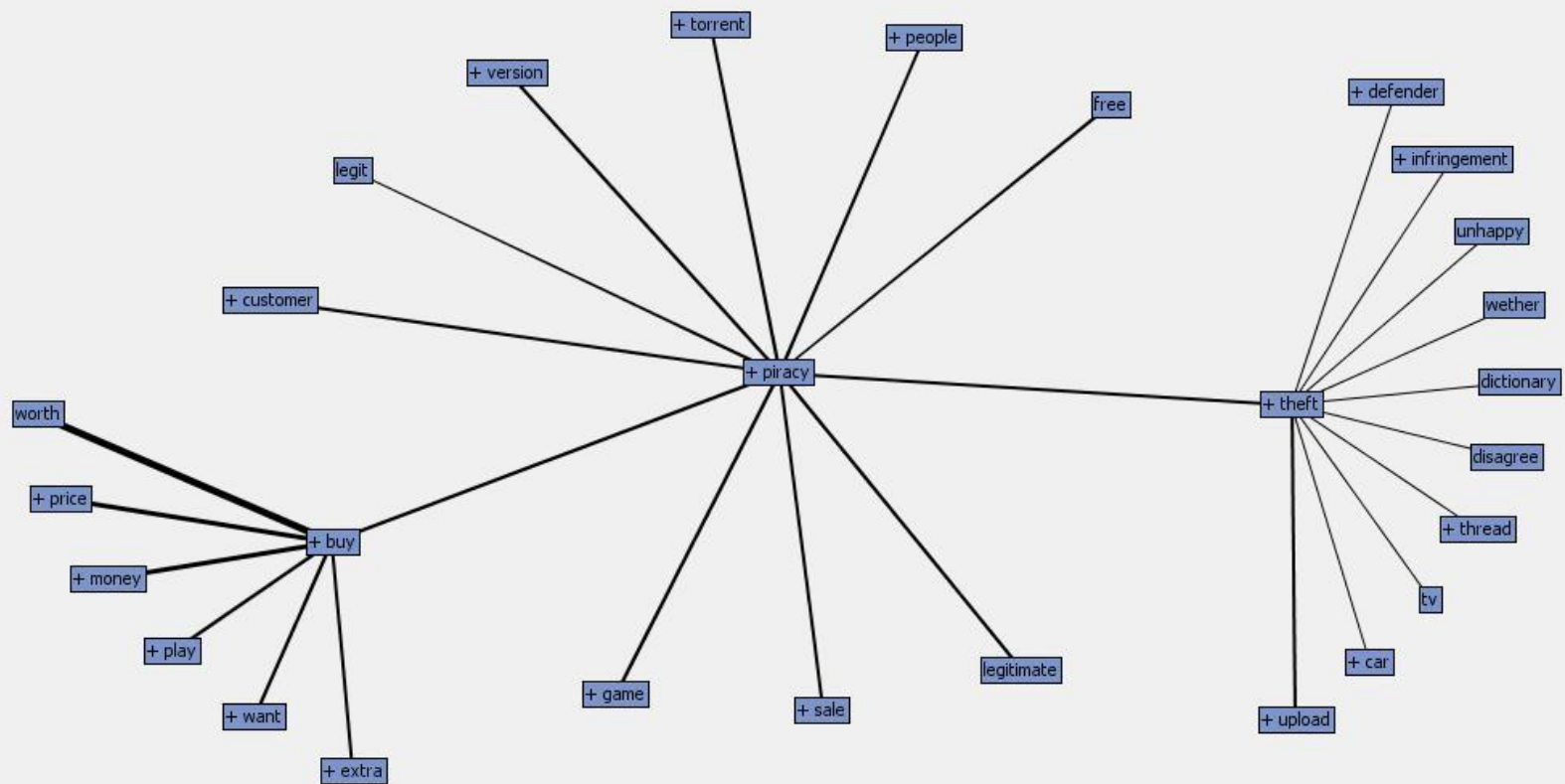
2.3. Piracy.

Na koniec przyjrzyjmy się otoczeniu słowa *piracy*. Na końcu pracy przedstawiony jest także wykres dla zbioru łączącego wszystkie trzy kategorie.

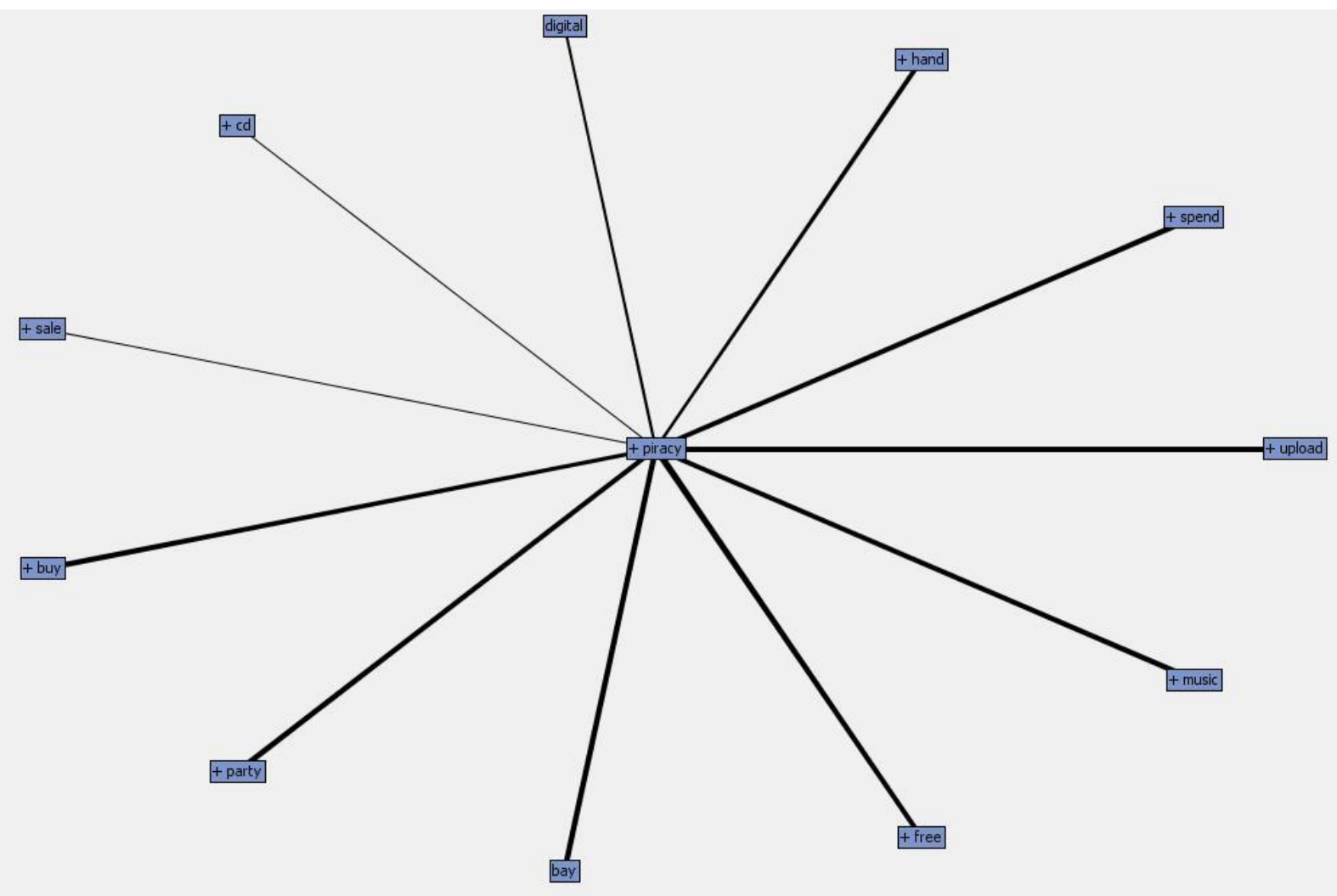


Rysunek 8. Concept links, tematyka: film, piracy.

Jak widać słowo *piracy* nieraz łączone jest z synonimami kradzieży. Widać także, że komentarze dotyczą i innych wymiarów piractwa – obok słów takich jak *file* (plik), mamy także nośniki fizyczne, jak np.: *dvd*, *cd*. Istotną różnicą między poszczególnymi branżami jest fakt, że w przypadku filmu nie pojawia się dyskusja o kupnie (zwroty *buy*), jak ma to miejsce w przypadku gier i muzyki. Może to sugerować, że filmy na DVD w większym stopniu narażone są na straty z powodu ‘piractwa’



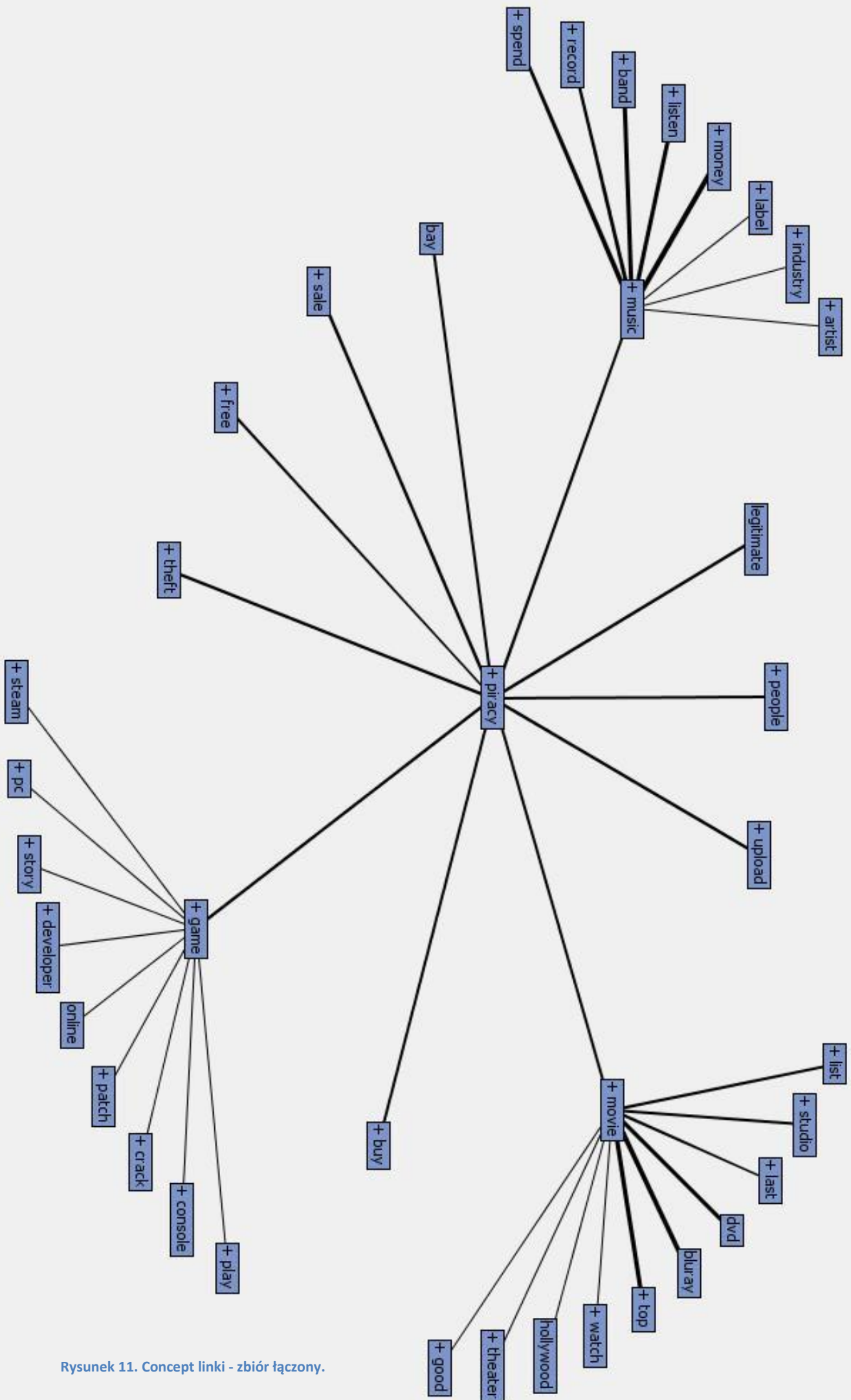
Rysunek 10. Concept links, tematyka: gry, piracy.



Rysunek 9. Concept links, tematyka: muzyka, piracy.

Na koniec wykonaliśmy także wykres *Concept Links* dla zbioru łązonego. Zamieszczony jest w rysunku 11.

Dla wykresu łązonego udało się uzyskać najczęstsze połączenia z poszczególnymi branżami: gier, filmów i muzyki. Pozwala to na zobaczenie na jednym wykresie, czego najczęściej dotyczą rozmowy w obrębie tych tematów. W przypadku gier, widzimy słowa związane z platformą (*pc, konsole*), crackami (*crack*), oraz wydawcami i dystrybutorami (*developer, steam*) i rozgrywką (*play, story, online*). W przypadku filmów, pojawiają się także rodzaje nośników/metody odbioru (*bluray, dvd, theater*), rolę wydawcy/dystrybutora spełnia tu *Hollywood* i *studio*. W przypadku przemysłu muzycznego ponownie obserwujemy wydawców/producentów (*industry, label*), ale też pojawiają się, co ciekawe, dwa nowe wątki, jakimi są pieniądze (*money, spend*), oraz rozmowy o samych artystach (*band, artist*). Prawdopodobnie w branży muzycznej istnieje silniejsza dyskusja na temat dystrybucji pieniędzy płaconych za sprzedawaną muzykę. W tym wypadku mamy małe zespoły autorów, którym pośredniczą wielkie firmy. W przypadku gier i filmów, w stworzeniu jednego produktu często uczestniczą setki osób.



Rysunek 11. Concept linki - zbiór łączony.

3. Podsumowanie

W niniejszej pracy przedstawiliśmy wyniki badania wypowiedzi dotyczących piractwa, wśród użytkowników popularnego serwisu TorrentFreak. Przeprowadziliśmy analizę dla wypowiedzi dotyczących trzech różnych tematów – piractwa gier komputerowych, piractwa muzyki oraz piractwa filmów. Dla danych z trzech zbiorów, odpowiadających trzem opisywanym tematom, przeprowadziliśmy analizę wykorzystywanych słów, ich połączeń oraz głównych grup tematycznych wypowiedzi. Analiza pokazała że wypowiedzi na trzy wybrane tematy wyraźnie różnią się od siebie – istnieją tematy, które poruszane są tylko w odniesieniu do konkretnej branży i żadnej innej.

Dzielenie się filmami zdaje się częściej być utożsamiane z rozprzestrzenianiem kultury na świecie. W tym wypadku też rzadziej zestawia się piractwo z kradzieżą. Temat piractwa muzyki łączy się z rozróżnieniem między wydawcami, a artystą. Może to mieć znaczenie poprzez rozkład wynagrodzenia między tymi podmiotami. Rzeczywiście też, to w przypadku muzyki dyskusje dotyczą także tematu pieniędzy i kupowania. Organizacje zbiorowego zarządzania, typu RIAA i MPAA są bardzo negatywnie postrzegane w środowiskach pirackich. Jest to relacja odwzajemniona, ponieważ obie instytucje silnie walczą z piractwem. W przypadku gier, wyłania się wątek zabezpieczeń oraz ich łamanie. Jest to cecha wyróżniająca tę branżę, w której wydawcy instalują skomplikowane mechanizmy mające uchronić gry przed ich piractwem. Wydaje się jednak, że owe zabezpieczenia są dość szybko łamane. Ich czołowym przedstawicielem w dyskusji jest serwis Steam, trudniący się w sprzedaży cyfrowej gier komputerowych.