ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 45, No. 12, pp. 1312-1318, 2018. 12
https://doi.org/10.5626/JOK.2018.45.12.1312

MOBA 게임 내 욕설 네트워크 분석을 통한 높은 영향력을 가진 악성 유저 탐지 방안

(Detection of Malicious Users with High Influence through Foul Language Network Analysis in MOBA Games)

안 동 현 † 김 휘 강 ††
(Dong hyun Ahn) (Huy kang Kim)

요 약 온라인 게임 산업의 발전과 더불어 게임 내 언어폭력은 심각한 사회적 문제로 떠오르고 있다. 하지만 단순한 필터링이나 신고제도로는 근본적인 문제를 해결하기 쉽지 않다. 따라서 본 논문에서는 욕설의 전파경향 분석과 언어폭력 중심유저 탐지를 위한 소셜 네트워크관점에서의 분석방법을 제안한다. 이 방법을 이용하여 전 세계적으로 인기를 끌고 있는 MOBA(Multiplayer Online Battle Arena)장르 게임인 DotA 2의 채팅로그 분석에 적용하였다. MOBA 게임의 경우, 하나의 큐(매치)에 속하는 유저가 제한되어있어 다른 장르의 게임보다 욕설 네트워크를 분석하기 좋은 플랫폼이다. 욕설을 남발하는 악성 유저의 경우 네트워크를 형성했을 때 높은 중심성(Centrality)을 갖는 경향이 있다. 이러한 특징을 이용하여 네트워크에서 욕설이 전파되는 경향을 파악하고 중심성(Centrality)이 높은 유저를 탐지하였다. 또한 해당 유저를 제재했을 때 전체 네트워크에 미치는 영향을 분석하였다. 본 논문에서 제안한 방법을 이용하면 욕설 사용으로 나쁜 영향을 미칠 수 있는 악성유저를 탐지할 수 있었다. 향후에는 유저들의 욕설 전파 유형을 분류하고 각 유형의 유저들이 갖는 특징을 분석한다.

키워드: 소셜 네트워크 분석, MOBA, 유저 행위 분석, 언어 폭력

Abstract In relation to the online game industry, verbal violence in the game has become a serious social problem. However, it is difficult to solve fundamental problems by simply filtering or using reporting systems. This study proposed a method to analyze the propagation tendency of the foul language and to detect malicious users in social network perspective. This method was applied to the analysis of the chat log of Defense of the Ancients 2(DotA 2), a popular MOBA(Multiplayer Online Battle Arena) genre game around the world. In the case of MOBA games, there are usually limited users belonging to one queue, which is a good platform for analyzing foul language networks as compared to other games. Verbally abusive malicious users tend to have high centrality when they form a network. Using these features, we analyzed the propagation tendency of the foul language on the network and detected users with high centrality. We also analyzed the effect on the whole network when the user was restricted. With the proposed method, we were able to detect malicious users who used the foul language. For future works, we will classify the spreading types in the foul language network and analyze users for each type.

Keywords: social network analysis, MOBA, user behavior analysis, verbal violence

† 학생회원 : 고려대학교 정보보호대학원 정보보호학과
jackie0304@korea.ac.kr
†† 종신회원 : 고려대학교 정보보호대학원 정보보호학과 교수(Korea Univ.)
cenda@korea.ac.kr
(Corresponding author임)

논문접수 : 2018년 7월 11일
(Received 11 July 2017)
논문수정 : 2018년 11월 1일
(Revised 1 November 2018)
심사완료 : 2018년 11월 2일
(Accepted 2 November 2018)

정보과학회논문지 제45권 제12호(2018. 12)

1. 서 론

온라인 게임은 웹과 더불어 가장 성공적인 서비스 중 하나이다. 인터넷의 발달과 함께 전 세계적으로 온라인 게임을 이용하는 사용자가 폭발적으로 증가하고 있고, 최근 출시된 국내 게임사인 블루홀의 배틀 그라운드의 경우 동시접속자가 300만 명을 넘어섰다. 이처럼 온라인 게임 산업이 계속해서 성장하고 있는 반면, 온라인 게임이 우리 사회에 미치는 부정적인 영향도 커지고 있다. 대부분의 온라인게임이 다중 사용자(multiplayer)를 기반으로 개발되기 때문에 게임 내 환경은 현대사회의 축소판이라고 할 수 있고, 실제로도 우리사회에서 일어날 수 있는 사회적 이슈들이 게임 내에서 일어나고 있다. 게임 중독과 더불어 가장 심각한 문제로 떠오르는 것이 게임 내 언어폭력문제이다. MMORPG(Massive Multiplayer Online Role Playing Game)가 유행하던 때부터 언어폭력문제가 불거져왔지만 최근 언어폭력이 심각한 수준에 다다랐다. 경찰청의 통계자료에 따르면 2014년 8,800여 건에 불과했던 사이버 명예훼손·모욕이 2015년에는 15,043건으로 전년도 대비 69.4% 증가했고, 그에 따라 온라인 게임 내 언어폭력 행위로 인해 처벌받는 사례도 증가하는 추세이다.

이와 같은 온라인 게임에서의 언어폭력 증가는 온라인의 익명성으로 인한 도덕적 판단능력 상실과 뇌과학적 근거로 뒷받침된다. 연구결과에 따르면 사람이 욕설을 들었을 때 뇌의 변연계 부분이 활성화된다. 이 부위는 먹기, 마시기, 불안, 공격성 등과 같은 인간의 가장 기본적인 욕구와 감정에 관여하고 기억과도 밀접한 관련이 있다. 욕설을 비롯한 막말은 인간의 가장 원초적인 부분을 자극하여 듣는 순간 당사자의 이성이 마비되고 감정에 휘둘려 화를 내며 대응하게 된다. 그렇기 때문에 깨끗한 채팅 환경을 위해서는 언어폭력의 중심에 위치하며 습관적으로 욕설을 하는 유저를 탐지하고 게임으로부터 배제하는 것이 필요하다.

현재 서비스 중인 대부분의 온라인 게임들의 경우에는 이러한 언어폭력 행위를 방지하기 위해 금지어 리스트에 포함된 단어들을 토대로 채팅창의 비속어를 자동으로 필터링하는 방식의 서비스를 제공하고, 추가적으로 사용자들의 신고를 통해 언어폭력을 행하는 유저들에게 게임상의 불이익을 주는 등의 방식을 채택하고 있다. 최근에는 언어폭력문제의 심각성을 인지하고 필터링이나 제재의 기준을 높이는 등의 조치를 취하고 있다. 하지만 필터링 방식의 경우 쉽게 우회가 가능하고 욕설이 아닌 경우에도 필터링이 되어 유저 간의 원활한 커뮤니케이션을 방해한다. 또한 지나치게 엄격한 기준을 적용함으로써 비교적 언어폭력으로 분류하기 힘든 유저까지 제재하여 유저들의 원성을 사는 경우가 빈번하게 발생한다. 그렇기 때문에 욕설전파의 중심에 위치하는 유저를 찾아 적절하고 효과적인 제재를 하는 것이 필요하다.

본 논문에서는 이러한 문제를 해결하고자 온라인 게임 채팅에 대한 소셜 네트워크 기반의 언어폭력 분석방안을 제안한다. 여러 온라인 게임 장르 중에서 MOBA (Multiplayer Online Battle Arena)장르의 게임을 분석에 사용하였다. MOBA는 액션, 롤플레잉 및 실시간 전략 게임을 융합한 것으로, 한 플레이어는 팀의 한 캐릭터를 담당하며, 캐릭터를 성장시키고 팀 구성원들과 함께 전략적으로 상대 팀의 주요 구조를 파괴하는 게임이다. 2장에서 네트워크를 통한 정보 확산과 소셜 네트워크에서의 중심성에 관한 기존 연구들을 알아보았다. 3장에서는 소셜 네트워크 분석을 통한 욕설전파 경향 분석 및 언어폭력 중심 유저 탐지방법을 제안한다. 4장에서는 MOBA 장르의 게임인 Defense of the Ancient 2(DotA 2)의 로그 데이터를 사용하여 제안한 방법을 평가하였으며, 5장에서 결론을 내렸다.

2. 관련 연구

소셜 네트워크 분석(SNA)은 현재 사회학, 통신 공학, 경제학 등 다양한 분야에서 폭넓게 연구되고 있는 분석 방법으로, 개인 간의 관계가 확산되어 사람들 사이의 연결된 네트워크인 사회 연결망을 분석하는 것을 말한다.

Bakshy[1]는 정보 확산에서 소셜 네트워크의 역할을 조사하였다. 정보에 노출된 사람들은 정보를 유포할 가능성이 훨씬 크고 노출되지 않은 사람보다 훨씬 빨리 정보를 유포한다. 또한, 약한 유대관계가 비교적 정보의 전파에 중요한 역할을 할 수 있음을 분석하였다. 그러나 해당 연구에서는 분석에 이용한 페이스북 외부에서의 커뮤니케이션에 대한 영향을 고려하지 않았다는 한계점이 있다. Gao 등[2]은 이메일 네트워크를 통한 바이러스 전파 특성을 이용하여 네트워크의 구조와 인간 역학이 바이러스 전파에 어떻게 영향을 주는지 실험하였다. 해당 연구에서는 2개의 실제 메일 네트워크에 대하여 실험을 수행했지만 좀 더 다양한 환경의 메일 시스템에 대한 실험이 추가로 필요할 것으로 보인다. Kimura 등[3]은 ICM(Independent Cascade Model) 기반의 대규모 소셜 네트워크에서 정보 확산에 영향을 미치는 노드들의 순위를 계산하는 새로운 방법을 제안하였다. 전파 확률이 낮은 경우에는 제안된 방법이 좋은 결과를 보였으나 전파 확률이 높은 경우에는 데이터의 종류에 따라 추가적인 검증이 필요하다. Yan 등[4]은 네트워크의 소셜 구조와 사용자 활동 패턴을 분석하여 온라인 소셜 네트워크에서의 멀웨어 전파 특성을 분석하고, 이를 통해 효과적인 방어 방법을 제안하였다. 하지만 해당 연구

에서 사용한 네트워크 노드들에 대한 가정들의 현실성이 비교적 떨어지는 한계점이 있다. Tang 등[5]은 노드의 topic 분포, 노드 간 유사성 및 네트워크 구조와 같은 정보를 포착하기 위해 TFG(Topical Factor Graph)를 제안하고, 대규모 네트워크에서 topic-level의 사회적 영향을 모델링하는 주제 유사성 전파(TAP)를 제안하였다. 해당 연구에서는 실험을 통해 기존 방법에 비해 효과적이라는 결과를 확인했지만 커뮤니티 탐지(community discovery)와 같은 다른 적용방안에 대해서는 추가적인 검증이 필요하다. Bonchi[6]는 데이터마이닝 관점에서 바이럴 마케팅을 위한 영향 최대화 문제를 중점으로 소셜 네트워크에서 영향이 전파되는 방식에 대한 조사를 수행하였다. 해당 연구에서는 여러 확산 모델들과 영향 최대화(influence maximization) 방법들에 대한 성능 비교 방법에 대한 기준의 부재를 한계로 지적하였다. Goyal 등[7]은 실제 소셜 네트워크 데이터를 이용하여 소셜 그래프와 사용자의 행동 로그를 추출하여 영향 모델을 생성하고 모델을 학습할 수 있는 알고리즘을 제안하였다. 또한, 실험을 통해 실제 소셜 네트워크에서도 진정한 영향이 있음을 보였다. 해당 연구에서 제안한 모델의 경우 유저(노드)가 행동을 취할 시간을 예측하는 기법을 개발하였지만 본래 입소문 마케팅(viral marketing)의 가정의 경우 엣지가 시간의 영향을 무시하고 단순히 라벨에 의해 전파 확률을 갖기 때문에 시간적 측면에서 입소문 마케팅을 공식화하고 해결하는 것이 필요할 것으로 보인다.

Tatsuya 등[8]은 네트워크에 속한 두 노드 간의 엣지를 해당 노드 쌍이 상호작용하기 전에 공유하는 정보의 양으로 표현하고 이를 통해 네트워크에서의 중심성(centrality)를 측정하였다. 또한, 중심성이 하나의 집단(group)에 미치는 영향을 분석하였다. 해당 연구에서는 비교적 소규모라고 할 수 있는 156명의 사람을 대상으로 실험을 수행하였기 때문에 온라인 소셜 네트워크상에서의 검증을 위해서는 실제 대규모 소셜 네트워크 환경에서의 실험이 필요하다. Jung 등[9]은 소셜 네트워크에서의 영향 최대화(influence maximization) 문제를 해결하기 위한 새로운 알고리즘을 제안하였다. 해당 알고리즘은 IC(Independent Cascade) 모델과 그 확장인 IC-N(Negative) 모델상에서 기존 알고리즘들에 비해 뛰어난 성능을 보임을 실험을 통해 확인하였다. 하지만 선형 임계 모델과 같은 다른 영향 전파(influence diffusion)모델에 대한 알고리즘 적용 가능성에 대해서는 추가적인 검증이 필요하다.

Woo 등[10-12]은 게임 내에 치팅(cheating) 플레이가 게임 내 소셜 네트워크를 통해 전파됨을 보였다. Kwak[13-15]의 연구들에서 보인 것과 마찬가지로 욕설뿐 아니라 게임 내 악성 행위들은 모두 게임 내 네트워크를 통해 전파되는 것을 밝혔다. 향후에는 온라인상에서의 치팅을 전파하는데 영향력이 큰 유저를 식별하고 확산 과정에서의 미치는 영향에 대하여 검증할 필요가 있다.

Kang 등[16]은 게임 내 악성 행위뿐만 아니라 호의적인 행위 역시 전파될 수 있음을 보였다. 게임 내 친구 관계 등 다양한 네트워크를 통하여 긍정적 행위가 전파됨을 확인할 수 있었다. 더불어 악성 행위가 유저들의 게임 이탈을 유발한다면, 호의적인 행위는 유저들의 게임 내 충성도를 높이게 되는 효과가 있음을 밝혔다. 해당 연구의 경우에도 하나의 게임에 대하여 실험을 진행했기 때문에 다른 게임 데이터 셋에서도 유사한 결과를 확인할 수 있는지에 대한 추가적인 실험이 필요하다.

3. 욕설 네트워크 중심성 기반 분석 방법론

3.1 욕설에 대한 기준

유저들 간의 채팅을 통해서 욕설 네트워크를 생성하기 위해서는 합리적인 기준을 통해서 각각의 유저가 입력한 채팅 내용에 대한 욕설 여부를 판단해야 한다. 실제 사회에서 사람들 간에 이야기하는 경우에는 문법에 어긋나는 문장을 사용하는 경우가 많지 않다. 하지만 온라인상에서 사람들이 채팅으로 대화를 할 때는 약어나 오타와 같이 채팅 내용에 대한 정확한 분석을 방해하는 요소들이 존재한다. DotA 2의 경우에는 우리나라 유저들보다는 외국에 거주하는 유저, 그중에서도 북미 유저들이 많은 비율을 차지하기 때문에 채팅 내용이 대부분이 영어로 이루어진다. 영어 욕설에는 한국어보다 비교적으로 약어로 이루어진 욕설이 많기 때문에 분석이 까다로울 수 있다. 또한, 단순히 욕설로 인식되는 단어의

표 1 AFINN 데이터 셋
Table 1 AFINN dataset

word	score
abandon	-2
abandoned	-2
abandons	-2
abhor	-3
abhorred	-3
abhorrent	-3

표 2 욕설 리스트 예시
Table 2 Example of Cursing word list

word	score
nigga	-5
damn	-5
shit	-5

개수만 세는 식의 방법으로는 'fucking brilliant'와 같이 의미상으로 긍정적인 경우에도 욕설로 분류하여 결과에 잘못된 영향을 미칠 수 있다.

선행 연구들을 살펴보면, 온라인상에서의 채팅 특징이 잘 나타나는 트위터(Twitter) 사용자들의 트윗에 대한 감정분석을 위해 유인가(valence) 점수를 사용하여 오타나 약어들이 포함된 문장에서도 적절한 분석이 가능하다는 것을 보였다[13]. 유인가 점수는 각각의 단어들이 갖는 정적 또는 부적인 정도를 나타낸 지표이다. 각 단어들의 유인가 점수들은 더 큰 말뭉치(Corpus)의 유인가 정도를 평가하는데 사용될 수 있다. 유인가 점수를 나타낸 데이터 셋에는 ANEW, SentiWordNet, AFINN 등 다양한 데이터가 존재한다. 하지만 이와 같은 데이터 셋들에는 욕설에 해당하는 단어들이 많이 포함되어 있지 않다. 본 논문에서는 3382개의 영 단어들에 대한 기존 유인가 점수를 포함하는 AFINN-165 데이터 셋(표 1)과 함께 1621가지 욕설이 포함된 욕설 리스트(표 2)를 사용하였다. 욕설 리스트에 포함된 단어의 경우 AFINN 데이터의 유인가 점수 범위인 -5 이상 5 이하의 값 중 가장 부적인 점수에 해당하는 -5를 부여하고 식 (1)을 사용하여 유저들의 채팅에 대한 유인가 점수를 계산하였다.

본 논문에서는 계산된 유인가 점수가 0.5보다 큰 경우 전체 채팅 중에서 부적이거나 욕설에 해당하는 단어의 비중이 큰 것으로 보고 욕설이라고 판단하였다.

$$w = \frac{PR \times 5 + \sum_{1}^{5} NC_i \times i}{PR \times 5 + \sum_{1}^{5} NC_i \times i + \sum_{1}^{5} NNC_i \times i} \quad (1)$$

w : 유인가 점수(valence score)
NC_i : Negative score가 i인 단어의 수
NNC_i : Non-negative score가 i인 단어의 수
PR : 욕설에 해당하는 단어의 수

3.2 욕설 네트워크 생성 방안

MOBA 장르의 게임의 경우에는 하나의 매치에 10명의 유저가 참여한다. 각 유저들은 5명씩 팀을 나눠서 상대방의 구조물을 먼저 파괴하는 쪽이 승리를 하는 방식으로 진행된다. 유저 개개인의 실력도 승리에 영향을 미치지만 단순히 한 명의 유저가 가진 실력이 다른 유저들에 비해 뛰어나다고 해서 매치에서 승리하기는 쉽지 않다. 팀을 이루는 유저들의 팀워크가 큰 영향을 미치고 다른 유저들의 플레이가 개개인의 승패와 직결되기 때문에 다른 장르의 게임들보다 유저들 간의 언쟁이 빈번하게 발생한다. 이러한 MOBA의 특징은 MMORPG와 같은 다른 장르의 게임들에 비해 비교적 채팅의 대상을 명확하게 지정할 수 있다.

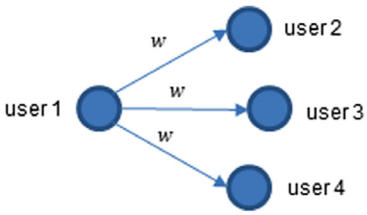

그림 1 가중치가 있는 방향 그래프
Fig. 1 Weighted directed graph

본 논문에서는 하나의 매치에 속한 유저들 간의 전체 채팅을 이용하여 채팅 네트워크를 생성한다. 각 유저들은 하나의 노드를 생성하고 한 유저가 플레이 중인 게임 내에서 채팅을 수행한 경우에 해당 매치에 속한 다른 유저들에게 방향성 있는 간선으로 연결한다. 간선의 비중(weight)은 유저가 매치 내에서 한 채팅 내용에 대하여 bag of words 기법을 적용한 후 채팅 내 포함된 단어들의 분포를 이용하여 식 (1)을 통해 계산된 유인가 점수를 할당한다(그림 1).

3.3 중심성 평가 척도

3.3.1 연결 중심성(Degree centrality)

연결 중심성은 가장 간단한 중심성 척도로써 한 노드에 연결된 간선의 개수로 중심성을 평가한다. 방향성이 있는 간선(directed edge)의 경우 들어오는 간선의 개수를 통해 노드의 인기도를 측정할 수 있고, 나가는 간선의 개수를 통해 그 노드의 영향력을 측정할 수 있다. 욕설 네트워크에서 연결 중심성이 크다는 것은 한 매치에 같이 매칭된 유저들로부터 욕을 듣거나 유저들에게 욕을 하는 빈도가 많다는 것을 의미한다. 즉, 들어오는 간선이 많다는 것은 다른 플레이어로부터 욕을 많이 듣는다는 것을 의미하고 나가는 간선이 많다는 것은 매치 내 다른 유저들에게 욕설을 많이 한다는 것을 의미한다. 이는 연결 중심성이 높은 유저가 한 매치 내에서 욕설을 전파할 뿐만 아니라 그 이후의 매치까지 욕설이 전파되는 데 영향을 미칠 확률이 높다는 것을 뜻한다. 따라서 게임 내 언어폭력 유저 모두를 제재하는 것이 불가능한 경우에 중심성이 낮은 유저들보다 욕설을 전파하는데 영향력이 큰 유저를 판별하여 제재하는 것이 효율적인 대안이 될 수 있기 때문에 연결 중심성을 평가 척도로 사용한다.

3.3.2 매개 중심성(Betweenness centrality)

매개 중심성은 노드들 간의 최단경로를 바탕으로 계산되는 수치로 네트워크에서 특정 노드의 중요성을 보기 위해서 해당 노드를 제외한 노드들 간의 경로(path)에서 해당 노드를 얼마나 거쳐서 가는지를 살펴보는 척도이다. 욕설 네트워크의 경우, 한 유저가 플레이 도중 욕설을 듣고 다음 매치에서 욕설을 하게 되면 해당 유

저로부터 다음 매치에 포함된 유저들에게 욕설이 전파된 것으로 볼 수 있고 이때 유저들 간에 경로가 형성된다. 욕설전파의 중심이 되는 유저일수록 특정 유저들 간의 경로에 포함될 확률이 높다. 즉, 매개 중심성이 높다는 것은 욕설 네트워크에서 욕설이 전파되는 경로 중에 특정 유저가 위치하는 빈도가 높다는 것을 의미하며, 이는 매개 중심성이 큰 유저일수록 하나의 매치에서 다음 매치로의 욕설전파에 미치는 영향이 크다고 볼 수 있는 근거가 된다. 따라서 매개 중심성을 욕설 중심 유저를 탐지하기 위한 척도로 사용한다.

4. 실험 및 평가

4.1 데이터 셋

본 논문에서는 DotA 2의 50000개 랭크 래더 매치리플레이를 API를 통해 파싱하여 생성한 오픈 데이터(표 3)를 사용하였다. DotA 2의 경우 옵션 중에 익명으로 플레이를 할 수 있는 기능이 있어 데이터 셋에 포함된 익명의 유저들은 구분이 불가능하다. 채팅 네트워크 생성 시에는 이러한 유저들은 제외하고 네트워크를 생성하였다. 또한, 욕설이 전파되는 경향을 분석하기 위해 플레이 횟수가 10회 이상인 유저들에 대해서만 네트워크를 생성하였다. 이때 생성된 간선 중에서 비중이 0.5보다 작거나 같은 경우 본 논문에서는 긍정적인 정도가 더 큰 것으로 분류하였기 때문에 채팅 네트워크에서 간선의 비중이 0.5보다 큰 값을 갖는 간선만 필터링하여 욕설 네트워크를 생성하였다.

표 3 채팅 로그 데이터 셋 필드
Table 3 Chat log data set field

Field	Description
match_id	Unique identifier of specific match
acc_id	Unique identifier of specific player
player_slot	Player slot(5 per team, total of 10 in match)
chat	Chat message
nickname	Nickname of specific player
start_time	Time match started

4.2 욕설 네트워크 생성

4.2.1 In-degree와 out-degree간의 상관관계

생성된 네트워크(그림 2)에서 매치가 진행됨에 따라 동적으로 변하는 노드로 들어오는 간선의 개수(in-degree)와 노드에서 나가는 간선의 개수(out-degree) 간의 상관관계를 분석하였다. 이때, 계산된 상관계수의 값이 약 0.85 정도의 높은 수치를 보였다. 이는 들어오는 채팅과 나가는 채팅 간의 강한 양의 상관관계가 있음을 의미하고, 다시 말해 유저가 욕설이 포함된 채팅을 받았을 때, 채팅으로 욕설을 할 확률이 크다는 것을 의미한다. 욕설 네트워크에서 t개의 매치 이후에 욕설이 전파된 모습은 그림 3과 같다.

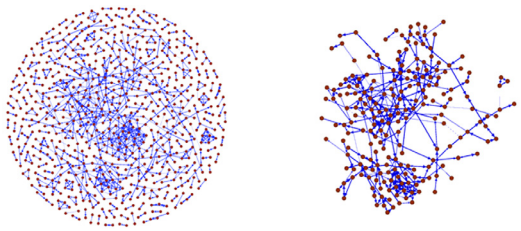

그림 2 욕설 네트워크와 최대 연결요소
Fig. 2 Foul language network and giant component

4.2.2 욕설 네트워크의 차수(degree) 분포

생성된 욕설 네트워크의 노드 차수 분포를 살펴보았을 때, 그림 4의 무작위 네트워크(random network)에서 나타나는 정규분포의 형태를 따르는 포아송 분포(Poisson distribution)가 아닌 척도 없는 네트워크(scale-free network)에서의 차수 분포인 멱 법칙에 따른 분포(power law distribution)를 나타냈다. 척도 없는 네트워크의 경우 다른 노드들에 비해 비교적 많은 간선을 가지는 허브(hub)가 존재하기 때문에 차수 분포를 나타냈을 때 꼬리가 길게 늘어지는 형태(long tail)를 띤다(그림 5). 척도 없는 네트워크 상의 SI(Suspicious Infectious) 모델의 면역전략의 경우, 단순한 무작위 면역전략(random immunization)보다 특정 노드에 대한

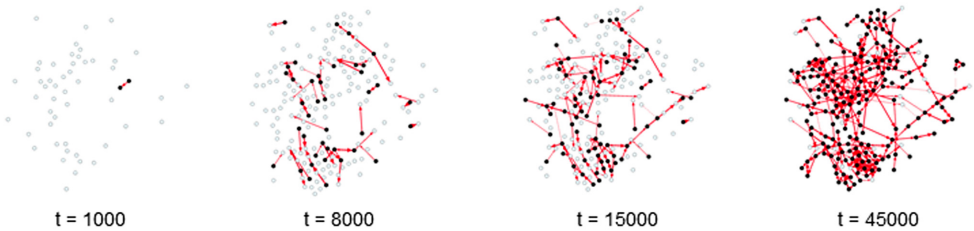

그림 3 t개 매치 이후의 욕설 전파
Fig. 3 Foul language propagation after t match

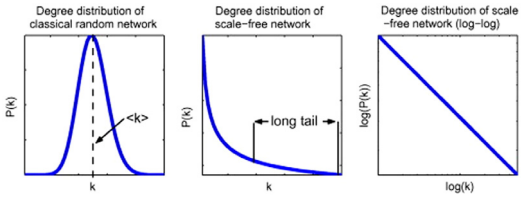

그림 4 네트워크 형태에 따른 차수 분포

Fig. 4 Degree(k) distribution of two types of network: random network with Poisson distribution(left), scale-free network with power law distribution (middle) and log-log plot of the power law distribution(right)

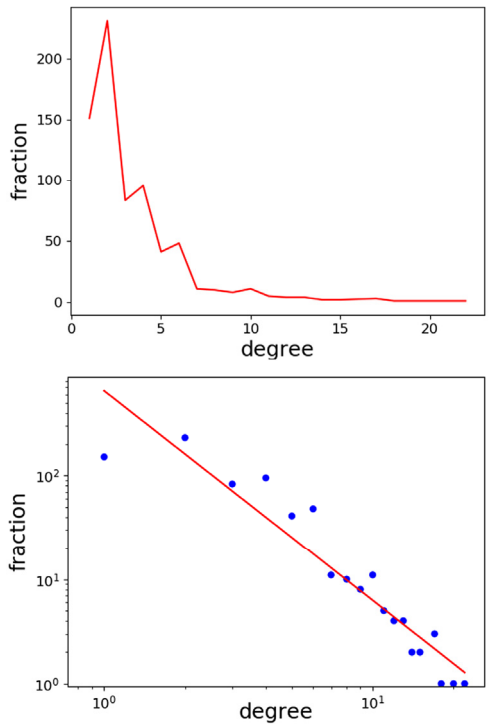

그림 5 욕설 네트워크의 차수 분포

Fig. 5 Degree distribution of foul language network(top) and log-log plot with regression line(bottom)

면역전략(target immunization)이 더 효과적이라는 연구결과가 존재한다[17]. 따라서 척도 없는 네트워크에서도 임의의 유저를 제재하는 것보다 전파의 중심이 되는 유저를 탐지하는 것이 효율적인 제재로 이어질 수 있음을 알 수 있다.

4.3 중심유저 탐지 및 제재

본 논문에서는 중심 유저 탐지를 위해 그래프 시각화 툴인 gephi에서 제공하는 통계적 기능을 사용하여 연결 중심성, 매개 중심성을 계산하고, 각 중심성 척도별로 상위 3개 노드를 선택하였다. 또한, 추가적으로 척도 없는 네트워크에서 허브의 역할을 하는 상위 3개 노드들도 선택하여 제재 효과를 비교하였다. 중심 유저의 제재 효과를 비교해 보기 위해서는 전체 욕설 네트워크에서 각 척도 별 추출된 노드를 제거했을 때 네트워크를 구성하는 노드들의 평균 차수의 변화와 연결요소 수의 변화를 살펴보았다. 평균 차수의 경우 전체 네트워크에서 욕설이 전파되는 빈도를 파악할 수 있고, 연결요소의 수는 특정 유저를 제재했을 때 연결요소의 수가 많아지는 경우 한 연결요소에서 다른 연결요소로의 욕설전파를 막을 수 있음을 의미하기 때문에 제재 효과를 비교하기 위한 평가 척도로 사용하였다.

표 4 평가 척도 별 상위 3개 노드 제거 시 네트워크 토폴로지 특징

Table 4 Topological features after eliminating top 3 nodes by evaluation scale

	Avg. Degree	# of connected component
Original	2.13	85
Degree centrality	1.94	92
Betweenness centrality	2.01	89
Hub	1.98	88
Random	2.11	84

실험 결과를 살펴보았을 때, 표 4에서 알 수 있듯이 욕설 중심 유저를 제재하기 위해 사용한 3가지 중심성 척도 모두 임의로 3개의 노드를 선택하여 제거한 경우보다 전체 네트워크의 평균 차수가 비교적 크게 감소하였고, 기존의 연결요소의 수보다 노드를 제거하고 난 뒤에 네트워크에서 떨어져 있는 연결요소의 수가 더 많아진 것을 알 수 있었다. 그중에서도 연결 중심성이 가장 높은 3개 노드를 제거했을 때 평균 차수의 감소가 가장 컸고, 연결요소의 증가도 가장 크게 나타났다.

5. 결 론

본 논문에서는 MOBA 장르 게임은 DotA 2의 채팅 로그 데이터를 이용하여 욕설 네트워크를 생성하고 해당 네트워크가 가지는 특징을 살펴보았다. 욕설 네트워크의 경우 네트워크의 차수 분포를 살펴보았을 때 무작위 네트워크가 아닌 척도 없는 네트워크의 특징을 가지는 것을 보였다. 이것을 바탕으로 욕설 네트워크에서의 욕설전파를 막기 위한 전략을 세우는 경우, 척도 없는 네트워크상의 SI 모델에서 면역전략을 세울 때 무작위 면역전략보다 특정 노드에 대한 면역전략이 더 효율적이라는 점을 이용하였다. 그 결과 욕설 제재 전략을 수립할 때 임의의 유저를 선택하는 것보다 욕설 전파에

영향력이 큰 특정 유저를 선택하여 제재하는 것이 효과적임을 보였다. 우선적으로 제재가 필요한 중심 유저를 선정할 때는 소셜 네트워크 분석방법을 이용하여 욕설 네트워크에서 높은 연결 중심성과 매개 중심성을 갖는 유저를 탐지할 경우 높은 제재 효과를 얻을 수 있을 것으로 기대하였다.

향후에는 유저들의 욕설 전파 유형을 분류하여 각 유형의 유저들이 갖는 특징에 대한 분석을 수행하고, 장기간의 데이터를 바탕으로 제재 이후 유저들의 욕설 성향에 대하여 연구를 진행한다.

References

[1] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," *Proc. of the 21st international conference on World Wide Web*, pp. 519-528, 2012.

[2] C. Gao, J. Liu, and N. Zhong, "Network immunization and virus propagation in email networks: experimental evaluation and analysis," *Knowledge and information systems*, Vol. 27, No. 2, pp. 253-279, 2011.

[3] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 259-271, 2016.

[4] G. Yan, G. Chen, S. Eidenbenz, and N. Ni, "Malware propagation in online social networks: nature, dynamics, and defense implications," *Proc. of the 6th ACM Symposium on Information, Computer and Communications Security*, pp. 196-206, Mar. 2011.

[5] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 807-816, 2009.

[6] F. Bonchi, "Influence Propagation in Social Networks: A Data Mining Perspective," *IEEE Intelligent Informatics Bulletin*, Vol. 12, pp. 8-16, 2016.

[7] Y. Goyal, F. Bonchi, and L.V. Lakshmanan, "Learning influence probabilities in social networks," *Proc. of the third ACM international conference on Web search and data mining*, pp. 241-250, Feb. 2010.

[8] T. Kameda, Y. Ohtsubo, and M. Takezawa, "Centrality in sociocognitive networks and social influence: An illustration in a group decision making context," *Journal of personality and social psychology*, Vol. 73, No. 2, 1997.

[9] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," *Data Mining(ICDM), 2012 IEEE 12th International Conference*, pp. 918-923, 2012.

[10] Woo, J., Kang, S. W., Kim, H. K., and Park, J., "Contagion of Cheating Behaviors in Online Social Networks," *IEEE Access*, 6, 29098-29108, 2018.

[11] Woo, J., Kang, A. R., and Kim, H. K., "The contagion of malicious behaviors in online games," *ACM SIGCOMM Computer Communication Review*, Vol. 43, No. 4, pp. 543-544 ACM, 2013.

[12] Ki, Y., Woo, J., and Kim, H. K., "Identifying spreaders of malicious behaviors in online games," *Proc. of the 23rd International Conference on World Wide Web*, pp. 315-316. ACM, 2014.

[13] J. Blackburn and H. Kwak, "Stfu noob!: predicting crowdsourced decisions on toxic behavior in online games," *Proc. of the 23rd international conference on World Wide Web*, pp. 877-888, Apr. 2014.

[14] Kwak, H., Blackburn, J., and Han, S., "Exploring cyberbullying and other toxic behavior in team competition online games," *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, pp. 3739-3748, 2015.

[15] Kwak, H., and Blackburn, J., "Linguistic analysis of toxic behavior in an online video game," *International Conference on Social Informatics*, Springer, Cham., pp. 209-217, Nov. 2014.

[16] Kang, A. R., Kim, H., Woo, J., Park, J., and Kim, H. K., "Altruism in games: Helping others help themselves," *Network and Systems Support for Games (NetGames), 2014 13th Annual Workshop on IEEE*, pp. 1-6, 2014.

[17] W. J. Bai, T. Zhou, and B. H. Wang, "Immunization of susceptible-infected model on scale-free networks," *Physica A: Statistical Mechanics and its Applications*, Vol. 384, No. 2, pp. 656-662, 2007.

안 동 현

2018년 건국대학교 소프트웨어학과 졸업(학사). 2018년~현재 고려대학교 정보보호대학원 재학(석사). 관심분야는 온라인 게임 보안, 자동차 보안, IoT 보안

김 휘 강

정보과학회논문지
제 45 권 제 11 호 참조