

Relationships Among User Group, Gender and Self-disclosure in Social Media

Phil-Sik Jang*

Abstract

In recent years the privacy issue on social media is often being discussed. The purpose of this study is to explore the relationships among user gender, user group according to user activity level (highly active vs less active) and self-disclosure in social media. We collected a total of 180 million tweets issued by 13 million twitter users for 12 months and investigated attributes of tweet (user's profile, profile image, description, geographic information, URL) which are related to self-disclosure and boundary impermeability. The results show there are significant ($p < 0.001$) interactions between user gender, user group and each attribute of tweet that are related to self-disclosure and show that the patterns of self-disclosure are different across attributes. The results also show that the mean self-disclosure scores and boundary impermeability of top 10% highly active users are significantly higher than other less active users for all genders.

▶ Keyword: Social media, User group, Gender, Self-disclosure, Twitter

I. Introduction

참여와 개방의 웹 2.0 시대와 함께 등장한 온라인 도구이며 플랫폼인 소셜 미디어(social media)는 사용자 하여금 웹상에서 자발적으로 생각과 정보 등을 공유하고 관계를 생성, 확장시킬 수 있도록 한다. 소셜 미디어는 사적영역과 공적영역이 중첩된 표현공간이라는 점[1]과 최근 급속한 보급이 이루어지고 있다는 점에서 사용자들의 개인 정보보호 문제가 주목 받고 있다[2]. 서로의 정보를 공개하는 자기노출(self-disclosure)은 사람 간의 의사소통에 있어 필수적인 요소이다[3]. 하지만 온라인 상에서 작동되는 소셜 미디어에서는 본인이 자발적으로 공개한 정보뿐만 아니라 의도하지 않은 정보까지도 실시간으로 광범위하게 확산될 수 있는데, 트위터(twitter)와 같은 개방형 소셜 미디어에서는 이러한 부정적 파급효과가 더 크게 나타나게 된다[4,5].

최근, 소셜 미디어 환경에서의 정보보호 및 자기노출이 사회적, 학문적으로 관심의 대상이 되면서, 자기노출 특성, 양상과 관련 영향요인들에 대한 연구들이 발표되고 있다. 즉, 사용자속한 국가, 언어 문화권에 따른 자기노출 양상에 대한 연구

[6-10], 미니 홈피 등에서의 자기노출 특성[11], 자기노출에 영향을 주는 태도[12], 동기요인[13] 및 집단 별 동질성 인식 [1] 등이 연구주제로 다루어지고 있다. 또한 사용자 성별에 따라 자기노출 특성이 어떤 차이를 보이는 가에 대한 연구들 [14-16]도 진행되었는데, 이들 연구 결과들은 아직 서로 상반되고 일관성 없는 결과들을 보이고 있다.

본 연구에서는 개방형 소셜 미디어인 트위터(twitter) 사용자들의 자기노출이 사용자의 성별 및 사용자 그룹과 어떤 관련이 있는지, 그 양상이 어떠한지를 분석하고자 하였다. 연구 방법론 측면에서, 선행 연구들은 대부분 연구자들에 의해 선택된 소수의 사용자만을 대상으로 하는 설문응답 방식을 이용하였다. 본 연구에서는 기존 설문응답 방식에 비해 전체적인 양상을 파악하는 데 유리할 것으로 기대되는[17] 대규모 온라인 데이터 수집 및 분석 방법을 적용하였다.

온라인 및 소셜 미디어 사용자들의 활동 패턴이 멱법칙(power law)을 따르는 것으로 알려져 있으나[18,19], 이를 고려한 연구들은 아직 찾아보기 힘들다. 본 연구에서는 집단 별

• First Author: Phil-Sik Jang, Corresponding Author: Phil-Sik Jang

*Phil-Sik Jang (philsjang@gmail.com), Dept. of IT & Logistics, Sehan University

• Received: 2018. 03. 21, Revised: 2018. 03. 31, Accepted: 2018. 04. 10.

• This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2014S1A5A2A03066219)

행동 패턴 즉, 사용자들을 활동(활성) 수준에 따라 분류하고 이들 집단에 따라 자기노출이 어떤 차이를 보이는 지를 살펴보고자 한다.

II. Theoretical Background

소셜 미디어를 이용한다는 것은 작성자가 공유하고자 하는 생각과 정보뿐만 아니라 작성자 자신의 정보까지 노출된다는 것을 의미한다[1]. 이러한 자기노출(self-disclosure)은 ‘자기 자신에 대해 밝히고 자신을 다른 사람에게 공개함으로써 다른 사람들이 자신을 알 수 있도록 하는 행위’로 규정된다[20]. 자기노출은 온라인(on-line)이든 오프라인(off-line)이든, 사람 간의 커뮤니케이션에 있어 중요한 역할을 하지만, 노출방식과 범위, 효과 등에 있어 차이를 보이는 것으로 알려져 있다[1]. 온라인에서는 자신에 대한 이미지를 오프라인에 비해 제약 없이 드러낼 수 있으며, 억제되었던 생각이나 느낌을 더 자유롭게 표현하는 경향이 있다[21]. 또한, 소셜 미디어와 같은 온라인 매체는 근본적으로 자기노출과 아주 밀접한 관계를 가진다. 상당수의 블로그와 SNS에서는 회원가입 시 일정 수준 이상으로 자기정보를 공개할 것을 요구하며, 이에 따라 블로그와 SNS를 이용한다는 것 자체가 자기정보를 자발적으로 온라인에 공개한다는 것을 의미한다. 그리고 이렇게 노출된 정보는 자신의 의도와 상관없이 대중에게 실시간으로 확산되기 때문에 오프라인과 비교할 수 없는 큰 파급효과를 가져올 수 있다. 이에 따라 소셜 미디어에서의 개인 정보보호와 자기노출은 서로 긴장관계에 놓이게 되며[21], Petronio[22]는 이러한 긴장관계를 커뮤니케이션 프라이버시 관리(CPM: Communication Privacy Management) 이론으로 설명한다. 이에 따르면, 개인의 자발적 의사에 의해 공개된 사적인 정보는 집단의 공동 자산(collectively owned)이 되며, 정보를 접한 사람들(집단) 내에서 일정한 프라이버시 규칙에 의해 관리된다. Petronio[22]는 이러한 프라이버시 규칙이 소속 집단의 문화에 영향을 받으며, 어떤 정보를 어느 정도 노출시켜도 되는가가 이러한 규칙에 따라 결정된다고 주장하였다.

몇몇 실증적 연구들 또한 이러한 문화적 영향을 뒷받침하고 있는데, 한국과 일본의 트위터 이용자들을 대상으로 한 연구[6]에서는 문화적 차이가 자기노출(사용자 지리정보 공개여부)과 소통방식에 영향을 준다고 보고하였다. 한국과 미국의 대학생 SNS 이용자들을 대상으로 실시한 연구[7]에서 자기노출과 자기표현(self-representation)이 문화권에 따라 유의한 차이가 있는 것으로 나타났다. 오스트레일리아와 사우디아라비아[8], 미국과 중국 두 개 문화권 간[9] 연구에서도 문화적 영향이 보고되었으며, 10개 언어 문화권을 대상으로 대규모 데이터 분석을 실시한 연구[10]에서도 이러한 경향성이 관찰되었다.

문화적 특성뿐 만 아니라, 오프라인 환경에서와는 다른 양상

을 보이는 온라인 환경에서의 자기노출에 대해 다양한 측면에서 연구가 진행되어오고 있다. 싸이월드와 같은 미니홈피를 대상으로 자기노출 양상[11] 및 자기노출에 영향을 주는 태도요인[12], 동기요인[13] 그리고 집단별 동질성 인식[1] 등이 연구 주제로 다루어져 왔다.

최근 성별에 따른 자기노출 연구도 이루어지고 있는데, 페이스북(Facebook)과 오프라인에서의 자기노출 성별 차이를 분석한 연구[14]에서는 페이스북과 오프라인 모두 여성이 남성에 비해 자기노출 수준이 높다고 보고하였다. 하지만, 국내 페이스북 사용자를 대상으로 한 연구[16]에서는 성별에 따른 유의한 차이를 발견하지 못하였으며, 러시아의 온라인 데이트 사이트 회원들을 대상으로 한 연구[15]에서는 자기노출관련 항목에 따라 다르게 나타나는 등 연구에 따라 상이한 결과를 보이고 있다.

이상에서 언급된 연구들은 대부분, 1,000명 이하의 동질적 소규모 집단을 대상으로 하는 설문방식을 이용하였다. 이러한 전통적인 설문방식 이외에 최근에는 빅데이터 분석방법론이 도입되면서 소셜 미디어연구에 있어 API(Application Programming Interface)를 활용한 데이터 수집 및 분석 방법론이 주목받고 있는데, 사용자의 자발적인 대규모 데이터분석이 가능하여, 여러 측면에서 기존의 설문방식의 연구와는 다른 관점을 제시 할 수 있을 것으로 기대되고 있다[17].

대규모 데이터를 활용한 소셜 미디어 연구에 있어 고려해야 할 요인으로 언급되고 있는 것이 사용자의 활성화수준(user activity)이다. 트위터와 같은 소셜 미디어에서는 상대적으로 적은 수의 고향성 사용자들(highly active users)의 데이터가 전체의 대부분을 차지하는 멱법칙(power law)을 따른다[18]. 이 때문에 사용자 집단 분류 시 10/90 또는 1/9/90의 비율을 이용하여야 하며, 트윗 패턴 등에 있어 집단 간에 많은 차이를 보이는데[19], 이러한 집단 간 차이를 고려한 연구들은 아직 찾아보기 힘들다. 본 연구에서는 사용자 활동(활성) 수준에 따라 10/90 비율로 사용자 집단을 분류하고 사용자 집단 및 성별과 자기노출관의 관련성을 조사하였다.

III. Method

본 연구에서는 소셜미디어 데이터 취합을 위해 대표적인 소셜네트워크 서비스이자 마이크로 블로그 서비스인 트위터(twitter)를 이용하였다. 트위터에서 제공하는 Streaming API[23]를 이용하여 실시간으로 트위터 메시지 데이터(tweet)를 수집하였으며, 데이터 저장과 분석에는 R version 3.4.3[24]과 오픈소스 DBMS인 MonetDB[25]를 이용하였다. 수집 기간은 2014년 11월에서 2015년 10월까지 총 1년, 수집된 총 메시지 수는 1억8천만 개이며, 이들 트윗을 게재한 사용자 별로 사용자 자기노출과 관련 있는 속성들을 취합하였다(Table 1).

취합된 트윗들을 게재한 영어권 사용자 총 수는 2천4백만 명으로, 사용자 성별 판단에는 Blevins & Mullen의 gender package[28]를 이용하였다. 사용자는 트위터 메시지 게재 수 상위 10% 사용자 집단(highly active user)과 하위 90% 사용자 집단(less active user)로 구분하여 분석하였다.

Table 1에서 제시한 각 속성들은 불리언(boolean)자료 형으로 코딩하였으며, 사용자 본인을 following하는 follower수와 본인이 following하는 사용자수(friend)는 정수형으로 취합하고, 다음 수식(1)을 이용하여 경계 불투과도를 산정[29]하였다. 경계불투과도는 자기노출 정보가 얼마나 많은 사람에게 노출 가능한가에 대한 상대적 비율을 나타낸다.

$$Boundary\ Impermability = \frac{N_{follower}}{N_{follower} + N_{friend}} \quad (1)$$

where

$N_{follower}$: The number of followers the account currently has.

N_{friend} : The number of users the account is following.

Table 1. Collected Tweet Attributes and User Information [26,27]

Tweet attribute	Description
follower_count	The number of followers this account currently has.
friends_count	The number of users this account is following.
default_profile	When true, indicates that the user has not altered their user profile.
default_profile_image	When true, indicates that the user has not uploaded their own avata.
user description	The user-defined UTF-8 string describing their account.
geo-enabled	When true, indicates that the user has enabled the possibility of geotagging.
url	A URL provided by the user in association with their profile.

IV. Results

취합된 트위터 사용자들 중 남녀 성별이 확실하게 분류된 인원은 총 1천3백만 명이며, 상위 10% 사용자들과 하위 90% 사용자별로 정리하면 Table 2와 같다.

Table 2. A Summary of Collected Twitter Users

	Female	Male	Sum
Highly active users	588,660	440,545	1,029,205
Less active users	6,304,361	6,243,393	12,547,754
Sum	6,893,021	6,683,938	13,576,959

1. Tweet Attributes

1.1 User's Profile

Table 1에서 제시한 트윗 항목별로 사용자의 성별, 사용자 집단과 자기노출 간의 연관관계를 분석하기위하여 계층적 로그선형 분석(hierarchical log linear analysis)을 실시하였다. 데이터 처리에는 R을 이용하였으며, 모형 설정 및 분석에는 IBM SPSS V23을 이용하였다. Table 3은 사용자 프로파일공개 여부와 성별, 사용자 집단의 2×2×2 로그선형 분석 결과이다. 즉, 포화모형(saturated)으로부터 후진제거법을 이용하여 모형을 최적화한 결과, 세 가지 요인사이에 유의수준 0.001에서 통계적으로 유의한 교호작용이 존재함을 보여주었다. Fig. 1은 사용자 집단 및 성별로 사용자 프로파일 공개여부를 비율로 나타낸 것인데, 트윗 수 상위 10%에 속하는 여성 사용자들의 프로파일 공개비율이 가장 높은(91.6%) 것으로 나타났으며, 트윗 수 하위 90%에 속하는 남성 사용자들의 공개비율이 가장 낮은(47.6%) 것으로 나타났다.

Table 3. Results of log linear analysis (Gender×Group×Self-disclosure of User's Profile)

If Deleted Simple Effect is	DF	L.R. χ^2 Change	Pr(> t)
Gender×Group×Disclosure	1	1602.74	0.000***

*p<.05 **p<.01 ***p<.001

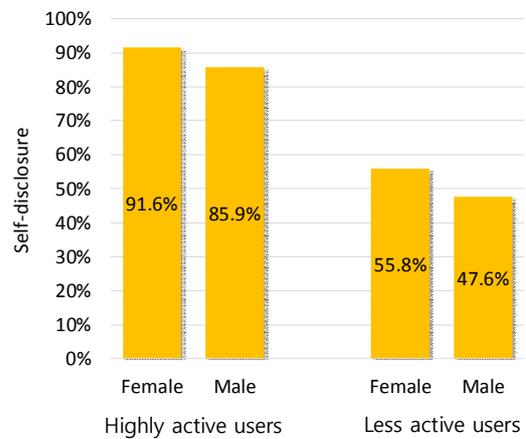


Fig. 1. Self-disclosure of User's Profile

1.2 User's Profile Image

Table 4는 사용자 프로파일 이미지 공개 여부와 성별, 사용자 집단의 2×2×2 로그선형 분석 결과로써, 후진제거법을 이용하여 모형을 최적화한 결과, 세 가지 요인사이에 유의수준 0.01에서 통계적으로 유의한 교호작용이 존재하는 것으로 나타났다.

사용자 집단 및 성별로 사용자 프로파일 이미지 공개비율을 나타내면 Fig. 2와 같은데, 사용자 집단과 성별에 상관없이 95%이상의 사용자들이 프로파일 이미지를 공개하는 것으로 나타났다. 큰 차이는 없지만 통계적으로 트윗 수 상위 10%에 속하는 여성 사용자들의 프로파일 공개비율이 가장 높고(99.7%), 트윗 수 하위 90%에 속하는 남성 사용자들의 공개비율이 가장 낮게(95.9%) 나타났다.

Table 4. Results of Log Linear Analysis (Gender×Group×Self-disclosure of User's Profile Image)

If Deleted Simple Effect is	DF	L.R. χ^2 Change	Pr(> t)
Gender×Group×Disclosure	1	17.65	0.006**

*p<.05 **p<.01 ***p<.001

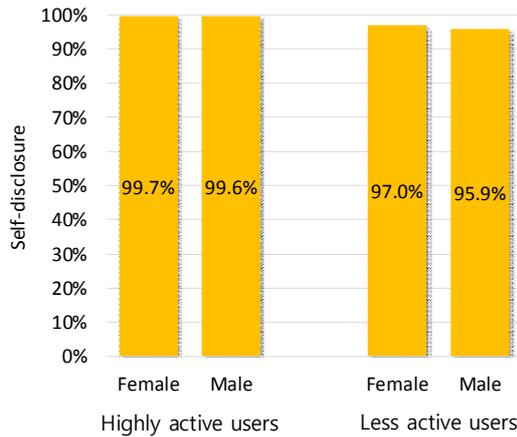


Fig. 2. Self-disclosure of User's Profile Image

1.3 User Description

사용자 설명 공개 여부와 성별, 사용자 집단의 2×2×2 포화 모형(saturated)으로부터 후진 제거법을 이용하여 모형을 최적화한 로그선형 분석 결과는 Table 5와 같다. 세 가지 요인 사이에 유의수준 0.001에서 통계적으로 유의한 교호작용이 존재하는 것으로 나타났는데, Fig. 3에서 볼 수 있는 것처럼, 트윗 수 상위 10%에 속하는 남성 사용자들의 사용자 설명 공개비율이 가장 높은(93.0%) 것으로 판단되며, 트윗 수 하위 90%에 속하는 남성 사용자들의 공개비율이 가장 낮은(47.6%) 것으로 나타났다.

1.4 User's Geographic Information

Table 6은 사용자의 지리정보 공개 여부와 성별, 사용자 집단의 2×2×2 로그선형 분석 결과로써, 세 가지 요인 사이에 유의수준 0.001에서 통계적으로 유의한 교호작용이 존재함을 보여준다. Fig. 4에서 볼 수 있는 것처럼, 3개 요인 간 교호작용은 통계적으로 유의하나, 성별 차이는 크지 않은 것으로 보이며, 트윗 수 상위 10%에 속하는 사용자들의 지리정보 비율이 트윗 수 하위 90%에 속하는 사용자들에 비해 상대적으로 높은 것으로 판단된다.

1.5 User's URL

사용자 URL 공개 여부와 성별, 사용자 집단의 2×2×2 로그선형 분석 결과는 Table 7과 같다. 세 가지 요인 사이에 유의수준 0.01에서 통계적으로 유의한 교호작용이 존재하는 것으로 나타났는데, 사용자들의 URL 공개 비율은 전체적으로 50% 미만인 것으로 보인다. 트윗 수 상위 10%에 속하는 남성 사용자들의 URL 공개비율이 상대적으로 가장 높았으며(41.5%), 트윗

수 하위 90%에 속하는 여성 사용자들의 공개비율이 가장 낮은(21.2%) 것으로 나타났다.

Table 5. Results of Log Linear Analysis (Gender×Group×Self-disclosure of User Description)

If Deleted Simple Effect is	DF	L.R. χ^2 Change	Pr(> t)
Gender×Group×Disclosure	1	2089.20	0.000**

*p<.05 **p<.01 ***p<.001

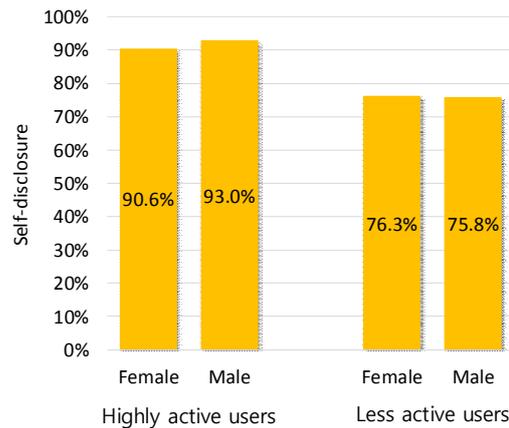


Fig. 3. Self-disclosure of User Description

Table 6. Results of Log Linear Analysis (Gender×Group×Self-disclosure of User's Geographic Information)

If Deleted Simple Effect is	DF	L.R. χ^2 Change	Pr(> t)
Gender×Group×Disclosure	1	30.10	0.000**

*p<.05 **p<.01 ***p<.001

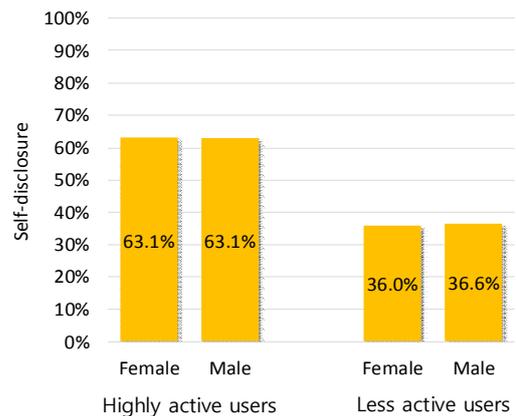


Fig. 4. Self-disclosure of User's Geographic Information

Table 7. Results of Log Linear Analysis (Gender×Group×Self-disclosure of User's Geographic Information)

If Deleted Simple Effect is	DF	L.R. χ^2 Change	Pr(> t)
Gender×Group×Disclosure	1	6.74	0.000**

*p<.05 **p<.01 ***p<.001

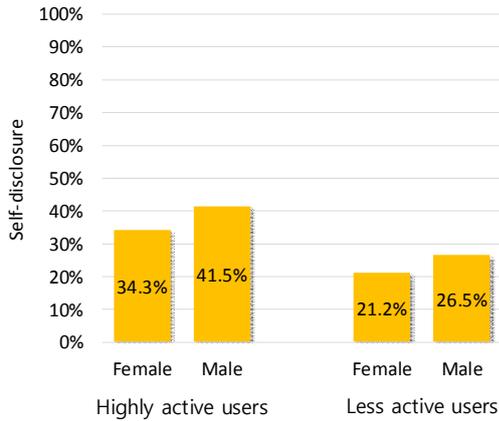


Fig. 5. Self-disclosure of User's URL

2. Self-disclosure Score

사용자 설명, 프로파일, 프로파일 이미지, 지리정보 및 사용자 URL 공개 여부를 합산 점수화(0~5점)하고, 이를 종속변수로 하는 분산분석을 실시하였다(Table 8).

분석결과, 유의 수준 0.001에서 사용자 집단, 성별 주 효과 및 교호작용이 통계적으로 유의한 것으로 나타났다. Fig. 6 은 사용자 집단과 성별에 따른 자기노출 합산점수 분포와 평균을 나타낸 것인데, 트윗 수 상위 10%에 속하는 남성 사용자들의 합산 평균(3.83)이 상대적으로 가장 높았으며, 트윗 수 하위 90%에 속하는 남성 사용자들의 합산 평균이 가장 낮은(2.82) 것으로 나타났다.

3. Boundary Impermeability

사용자별 follower 수와 friend 수를 이용하여 계산된(수식(1)) 경계불투과도를 종속변수로 하고, 성별과 사용자 집단을 독립변수로 실시한 분산분석(ANOVA)결과는 Table 9와 같다. 유의수준 0.001에서 주효과(성별, 사용자 집단) 및 교호작용이 통계적으로 유의한 것으로 나타났다. Fig. 7은 성별, 사용자 집단 별 경계 불투과도의 분포 및 평균값을 도시한 것인데, 트윗 수 상위 10%에 속하는 여성 사용자들의 경계불투과도 평균(0.61)이 상대적으로 가장 높았으며, 트윗 수 하위 90%에 속하는 남성 사용자들의 불투과도 평균이 가장 낮은(0.42) 것으로 파악되었다. 통계적으로 교호작용이 유의한 차이를 보이는 하나, 남녀 성별의 효과보다는 사용자 집단에 따라 경계 불투과도가 더 큰 영향을 받는 것으로 사료된다.

Table 8. Results of ANOVA (Dependant Variable: Self-disclosure Score)

	Df	SS	MS	F value	Pr(> t)
Group	1	887,454	887,454	626,506.4	0.000**
Gender	1	3,797	3,797	2,680.2	0.000**
Group×Gender	1	1,348	1,348	951.9	0.000**
Residuals	1.3e+7	1.9e+7	1		

*p<.05 **p<.01 ***p<.001

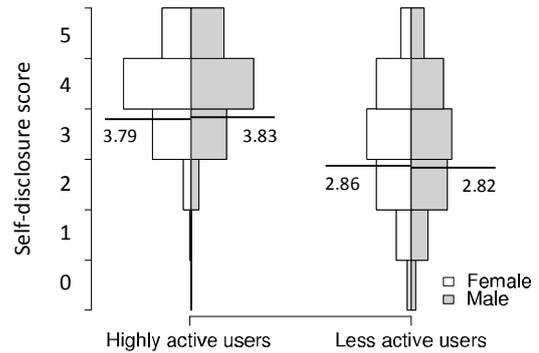


Fig. 6. Self-disclosure Score(group×gender)

Table 9. Results of ANOVA (Dependant Variable: Boundary impermeability)

	Df	SS	MS	F value	Pr(> t)
Group	1	887,454	887,454	626,506.4	0.000**
Gender	1	3,797	3,797	2,680.2	0.000**
Group×Gender	1	1,348	1,348	951.9	0.000**
Residuals	1.3e+7	1.9e+7	1		

*p<.05 **p<.01 ***p<.001

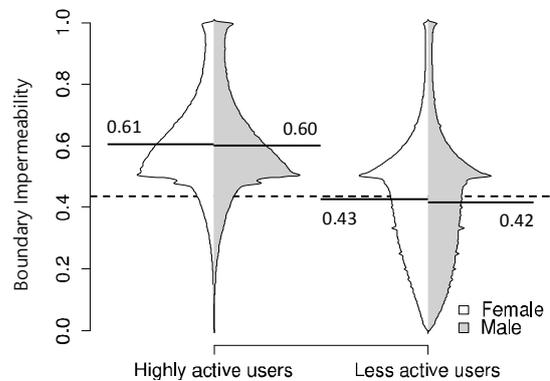


Fig. 7. Boundary Impermeability(group×gender)

V. Conclusions

본 연구에서는 개방형 소셜 미디어 사용자들의 성별, 사용자 집단(트윗 게시 상위 10%-하위 90%)과 자기노출이 어떤 연관 관계가 있는지를 정량적으로 파악하고자 하였다. 이를 위해 1년에 걸쳐 1억8천만 개 트윗을 취합하고 이들 트윗을 게시한 영어권 사용자 1천3백만 명을 대상으로 사용자 프로파일, 프로파일 이미지, 지리정보 및 사용자 정보, URL 공개 여부를 분석하였다. 이와 함께, 사용자별 follower 수 및 friend 수를 이용하여 경계불투과도를 분석하였다.

항목별 공개 여부와 성별, 사용자 집단의 2×2×2 로그선형

분석 결과, 모든 항목에 있어서 세 가지 요인사이에 유의수준 0.001에서 통계적으로 유의한 교호작용이 존재하는 것으로 나타났다. 항목에 따라 성별, 집단별 공개비율 추세가 다르게 나타났다는데, 일반적으로 트윗 수 상위 10%에 속하는 사용자들의 프로필 공개비율이 트윗 수 하위 90%에 속하는 사용자들의 공개비율이 보다 높은 것으로 파악되며, 통계적으로는 유의한 교호작용이 존재하나, 성별에 의한 차이는 크지 않은 것으로 사료된다.

5개 항목 합산점수를 종속변수로 하는 분산분석 결과 유의수준 0.001에서 성별, 사용자 집단 간 주효과, 교호작용이 모두 통계적으로 유의한 것으로 나타났다. 트윗 수 상위 10%에 속하는 남성 사용자들의 합산 평균이 상대적으로 가장 높았으며, 트윗 수 하위 90%에 속하는 남성 사용자들의 합산 평균이 가장 낮은 것으로 나타났다.

경계불투과도를 종속변수로 하는 분산분석 결과 또한, 유의수준 0.001에서 주효과(성별, 사용자 집단) 및 교호작용이 통계적으로 유의한 것으로 나타났다. 통계적으로 교호작용이 유의한 차이를 보이기는 하나, 남녀 성별의 효과보다는 사용자 집단에 따라 경계 불투과도가 더 큰 영향을 받는 것으로 사료된다. 즉, 사용자 정보의 노출 가능성은 성별에 따라 통계적으로 유의하나, 큰 차이를 보이지는 않는 것으로 생각되며, 트윗 수 상위 10%의 고향성 사용자 집단이 다른 집단에 비해 노출 가능성이 높은 것으로 판단된다.

본 연구는 개방형 소셜 미디어인 트위터에 한정되어 페이스북 등 폐쇄형 소셜 미디어 등을 포함한 다양한 측면에서의 분석이 이루어지지 못하였으며, 최근 데이터가 포함되지 못한 한계점을 가지고 있다. 하지만 본 연구는 대규모 데이터를 기반으로 개방형 소셜 미디어에서의 자기노출에 대한 실증적 이해를 증진할 수 있는 자료로 활용 가능할 것으로 기대된다. 차후, 다양한 SNS 사례 데이터 분석을 추가하고, 최근 데이터를 보완하여 장기간에 걸친 분석이 이루어진다면 더 의미 있는 결과가 도출될 것으로 판단된다.

REFERENCES

- [1] B. S. Chon, "Perceived Homophily by the Degree of Self Disclosure among SNS Users", Journal of the Korea contents association, Vol. 12, No. 5, pp. 170-178, 2012.
- [2] <http://www.internetsafety101.org/snsdangers.htm>
- [3] N. Collins and L. Miller, "Self-disclosure and Liking: a meta-analytic review," Psychological Bulletin, Vol.116, No.3, pp.457-75, 1994.
- [4] S. Utz and N. Kramer, "The Privacy Paradox on Social Network Sites Revisited: The Role of Individual Characteristics and Group Norms," Cyberpsychology: Journal of Psychosocial Research on Cyberspace, Vol.3, No.2, article 1, 2009.
- [5] M. A. Moreno, N. C. Fost, and D. A. Christakis, "Research Ethics in The MySpace Era," Pediatrics, Vol.121, No.1, pp.157-161, 2008, doi:10.1542/peds.2007-3015. PMID 18166570
- [6] S. E. Cho and H. W. Park, "Cross-National Comparison of Twitter Use between South Korea and Japan (An Exploratory Study)," International Journal of Contents, Vol.8, No.4, pp.50-55, 2012.
- [7] J. J. Yoo, D. Kim and J. Moon, "Exploring Cross-Cultural Differences in Self-Presentation and Self-Disclosure in Social Networking Sites: A Comparison of Korean and American SNS Users," Journal of Advertising and Promotion Research, Vol.1, No.2, pp.77-118, 2012.
- [8] H. A. Almakrami, Online self-disclosure across cultures: A study of Facebook use in Saudi Arabia and Australia, Ph.D. Thesis, Queensland University of Technology, 2015.
- [9] S. Luo, Cross-cultural differences between American and Chinese college students on self-disclosure on social media, Graduate Theses and Dissertations, pp.1-70, Iowa State University, 2014.
- [10] P. S. Jang, "Self-Disclosure and Boundary Impermeability among Languages of Twitter Users", Journal of the Korea contents association, Vol. 16, No. 4, pp.434-441, 2016.
- [11] Y. R. Kim, "Dialogical Writing in Virtual Space", Journal of Speech, Media and Communication Research), Vol. 3, No. 3, pp.31-57, 2004.
- [12] S. J. Lee and Y. B. Yeon, "Research Examining College Students' Self-Disclosure on Mini-Homepage", Korean Journal of Journalism & Communication Studies, Vol. 52, No. 3, pp.98-121, 2008.
- [13] Y. Choi. and S. H. Park, "The Effects of Social Media Usage on Social Capital", Korean Journal of Broadcasting, Vol. 25, No. 2, pp.241-276, 2011.
- [14] P. Sheldon, "Examining gender differences in self-disclosure on Facebook versus face-to-face." The Journal of Social Media in Society Vol 2, No. 1, 2013.
- [15] S. Kisilevich and M. Last, Exploring gender differences in member profiles of an online dating site across 35 countries. In Analysis of Social Media and Ubiquitous Data pp.57-78. Springer, Berlin, Heidelberg, 2011.
- [16] E. Y. Na, "Twitter Users' Online Self-Disclosure : Impacts of Risk Reduction and Benefit Expectancy", Korean Society For Journalism And Communication Studies, Vol. 57, No. 4, pp.124-148, 2013.
- [17] J. H. Lee, Ten faces of SNS, Communications books, 2013.

- [18] C. Anderson, *The long tail: Why the future of business is selling less of more*, New York, NY: Hyperion, 2006.
- [19] S. J. Tedjamulia, D. L. Dean, D. R. Olsen and C. C. Albrecht, *Motivating content contributions to online communities: Toward a more comprehensive theory*. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05)*, pp. 193b, Washington, DC: IEEE, 2005.
- [20] S. N. Jourard and P. Lasakow, "Some factors in self-disclosure," *Journal of Abnormal and Social Psychology*, Vol. 56, No. 1, pp.91-98, 1958.
- [21] H. J. Yoon, "Perceived Anonymity in an Online Social Support Community : A Hierarchical Structure of the Concept and its Effects on Online Public Disclosure", *Korean Journal of Journalism & Communication Studies*, Vol. 50, No. 6, pp. 305-332, 2006.
- [22] S. Petronio, *Boundaries of privacy: Dialects of disclosure*, Albany, NY: State University of New York Press, 2002.
- [23] <https://dev.twitter.com/streaming/overview>
- [24] <https://www.R-project.org/R>, Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017.
- [25] <https://www.monetdb.org/>
- [26] <https://dev.twitter.com/overview/api/users>
- [27] <https://dev.twitter.com/overview/api/tweets>
- [28] Blevins, C., & Mullen, L., "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction", *DHQ: Digital Humanities Quarterly*, Vol. 9, No, 3, 2015.
- [29] S. C. Walton and R. E. Rice, "Mediated disclosure on Twitter: The roles of gender and identity in boundary impermeability, valence, disclosure, and stage," *Computers in Human Behavior*, Vol.29, pp.1465-1474, 2013.

Authors



Phil Sik Jang received the B.E. degree in Naval Architecture from Seoul National University in 1990 and received M.S. and Ph.D. degrees in Industrial Engineering from KAIST in 1992 and 1998, respectively. Dr. Jang joined the faculty of the School

of Computer Science at Daebul University, Korea, in 1997. He is currently a Professor in the Department of IT & Logistics, Sehan University. He is interested in HCI, VR, AR and Big data analysis.