

페이스북 사용자 계정의 신뢰도 평가에 대한 연구*

박 정 은,^{1*} 박 민 수,¹ 김 승 주^{2*}¹고려대학교 정보보호연구원, ²고려대학교 사이버국방학과/정보보호연구원

The Reliability Evaluation of User Account on Facebook*

Jeongeun Park,^{1*} Minsu Park,¹ Seungjoo Kim^{2*}¹Center for Information Security Technologies(CIST), Korea University,²Department of Cyber Defense/Center for Information Security
Technologies(CIST), Korea University

요 약

스마트폰, 태블릿 PC 등 스마트 기기가 보급되면서 SNS(Social Network Services)의 사용자가 증가하고 있다. SNS는 연령이나 직업 등 사용자의 배경과 무관하게 많은 사람들과 인맥 형성을 할 수 있고, 정보를 쉽고 빠르게 전달하는 도구로 사용되고 있다. 최근에는 SNS의 장점을 악용하는 역기능이 이슈화되고 있다. 예를 들면, 계정 생성이 간편하고 인증절차 없이 가입할 수 있는 점을 통해 한 사용자가 다수의 계정을 생성하여 광고성 글 게재, 악성 댓글 등 악성 행위를 하거나, 친구관계를 맺어 사용자의 개인정보를 유출하는 것을 꼽을 수 있다. 본 논문에서는 세계적으로 가장 점유율이 높은 SNS인 페이스북을 연구대상으로 정한다. 따라서 페이스북 사용자들로부터 생성되는 정보를 안전하게 공유하고 서로 간의 신뢰관계를 보장하며, 악성 행위를 하는 사용자나 봇을 판단하기 위해 사용자의 행동을 기반으로 신뢰기준을 선정하여 이에 따른 신뢰점수를 사용자 계정에 적용시키는 방법을 제안한다.

ABSTRACT

Most people are connected to Social Network Services (SNS) through smart devices. Social Network Services are tools that transport information fast and easily. It does not care where he or she comes from. A lot of information circulates and is shared on Social Network Services. but Social Network Services faults are magnified and becoming a serious issue. For instance, malicious users generate multiple IDs easily on Facebook and he can use personal information of others on purpose, because most people tend to undoubtedly accept friend requests. In this paper, we have specified research scope to Facebook, which is one of most popular Social Network Services in the world. We propose a way of minimizing the number of malicious actions on Facebook from malignant users and malicious bots by setting criteria and applying reputation system.

Keywords: SNS, Facebook, Trust, Trust Relationship, Reputation System

접수일(2013년 6월 19일), 수정일(1차: 2013년 7월 26일, 2차: 2013년 10월 7일, 3차: 2013년 10월27일), 게재확정일(2013년 10월 28일)

* 본 연구는 방위사업청과 국방과학연구소의 지원(계약 번호 UD130007DD)과 미래창조과학부 및 정보통신산업진흥

원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2013-H0301-13-1003).

† 주저자, planet_in@korea.ac.kr

‡ 교신저자, skim71@korea.ac.kr(Corresponding author)

I. 서 론

SNS는 Social Network Services의 줄임말로 온라인상에서 사회 네트워크를 제공해주는 서비스를 의미하며, OSN(Online Social Network)라고도 한다. SNS는 1997년에 정의되었으며, 1998년 식스디그리스닷컴(Sixdegrees.com)이 런칭되면서 사용자의 프로필을 생성, 그들의 친구를 목록화 시키는 것으로 시작되었다(1). SNS는 온라인에서 불특정 다수와 관계를 형성할 수 있는 소셜미디어로 정의되기도 한다. 소셜미디어는 SNS에 가입한 사용자들이 서로 정보와 의견을 공유하면서 대인관계를 넓힐 수 있는 플랫폼이다.

SNS는 시간과 공간의 제약 없이 사용자간의 자유로운 의사소통이 가능하여 다양한 계층 간의 커뮤니케이션이 활발하게 이루어지고, 방대한 양의 정보가 공유될 수 있는 장점을 가지고 있다. 2012년 10월, 리서치 전문 기업인 '모니터링 닷컴'에서 발표된 통계자료에 따르면, 페이스북의 월간 실사용자는 10억명 이상이며, 평균 사용자의 친구 수는 130명으로 나타났다. 이는 미국에 한정된 결과가 아닌 인도, 프랑스, 영국 등 세계 각국의 통계치를 산정한 결과이다(2). 페이스북은 현재 국내 점유율도 1위를 차지하고 있으며, 다음 Fig.1.은 2013년 9월부터 10월 3일까지의 국내 SNS 점유율을 나타낸다(3).



Fig. 1. SNS Domestic Share

SNS의 장점을 토대로 사용자가 점점 증가하고 있지만, 부정적인 영향도 증가하는 추세이다. 예를 들면, 콘텐츠를 쉽게 게시할 수 있는 점을 이용하여 선정적인 개인 방송 동영상과 스팸을 유포되거나, 익명성으로 인한 도덕 불감증, 정보 왜곡으로 인한 루머 확산을 꼽을 수 있다. 또한, 페이스북 사용자들을 속여 신용사기 피해를 입히는 사례도 증가하고 있다(4). 이외에도 SNS는 누구나 쉽게 다수의 계정을 생성할 수 있어 문제가 되고 있다. 일반적으로 국내의 프로필 기반 SNS인 싸이월드와 웹 카페, 블로그의 경우에

회원 가입과 승인이라는 단계적인 본인 확인 절차를 거치게 된다. 하지만 최근에는 개인정보 수집을 최소화하기 위해 국내외 여러 SNS에서 고유식별정보를 요구하지 않으므로 이메일 인증, 휴대폰 인증, CAPCHA 등을 사용하고 있다. 이와 같은 인증 방법을 무력화하는 공격이 현재 많이 제시되고 있어 효과적인 본인 인증절차가 미비한 실정이다. 즉, 다양한 방법을 통해 악성 계정들이 생성되고 있으며, 이들은 불법적인 행위나, 음란성 글 게재, 악성 스크립트 삽입 등의 행위를 하여 정상 사용자에게 불편함을 주기도 한다. 그리고 SNS에서 친구 요청 시 사용자의 친구 수락여부를 위한 기준이 없기 때문에 현실에서 모르는 사람으로부터 친구 요청이 와서 이를 수락했을 때, 친구 요청자가 악의적으로 요청한 사용자의 개인정보를 획득하는 경우도 빈번히 일어나고 있다(5). 이러한 사례는 사용자가 친구 요청을 쉽게 수락하는 문제에서 비롯되는 것이다. 통계자료에 의하면, 페이스북에서 많은 사람들이 친구 요청 시 요청자가 누구인지 판단하지 않고 수락하는 경우가 대부분인 것으로 나타났다. 지역과 측정시기에 따라 다르지만, 평균적으로 요청 시 수락은 75%, 무시 및 거절은 25%로 확인할 수 있다(6-9). 이외에도 웹 기반의 봇넷을 SNS에 적용하여, 사용자의 개인정보를 획득하는 공격을 시도한 연구에 따르면, 공격 성공 가능성이 80%에 달하는 것으로 알 수 있다(10). 이렇듯 SNS의 장점을 악용한 악성 행위는 많은 위협을 내포하고 있다. 따라서 악성 행위를 하는 사용자와 악성 봇을 판단할 수 있는 방법이 필요하다.

본 논문에서는 연구 범위를 대표적인 프로필기반의 SNS인 페이스북으로 선정한다. 선정 이유는 현재 페이스북의 점유율이 국내뿐만 아니라, 국외에서도 가장 높고 나타나고 있으므로 제안 방법이 타 SNS보다 더 많은 의미를 가질 수 있기 때문이다. 따라서 페이스북에서 정보의 안전한 공유, 사용자 계정 간의 신뢰 보장, 그리고 악성 계정을 판단하기 위해 발생 가능성이 있는 보안 위협을 도출하고, 이를 최소화하기 위해

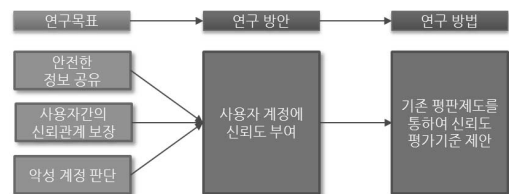


Fig. 2. The Proposal Method & Research goal

계정을 판단할 수 있는 신뢰도 평가 기준을 제안한다. Fig. 2.는 논문에서 제안하는 방법을 도식화 하였다.

논문의 구성은 2장에서 SNS의 유형을 정의하고 페이스북의 보안 위협을 도출한다. 또한, 신뢰와 평판 제도 정의 및 사례를 설명한다. 3장에서는 신뢰를 적용한 SNS 등 관련 연구에 대해 설명한다. 4장에서는 2장의 배경지식과 3장의 관련 연구를 바탕으로 본 연구에서 제안하는 방법을 제시하고, 5장에서는 제안하는 방법과 기존의 연구를 비교 및 검증한다. 마지막으로 6장에서 결론을 내린다.

II. 배경 지식

2.1 SNS의 유형과 페이스북

SNS는 사용자 간의 자유로운 의사소통과 정보 공유, 그리고 인맥 형성 및 확대를 도와주는 온라인 플랫폼이다. 온라인 환경인 SNS에서 사람들은 새로운 형식의 사회적 관계를 맺고 있으며, 이는 기존의 오프라인 환경에서의 사회적 관계와 다르게 성별, 나이, 직업, 지역 등의 제약 없이 다양한 배경을 가진 사람들과 자유롭게 관계를 형성하고 있다. SNS는 여러 가지 유형이 존재한다. Table 1.은 SNS의 기능에 따른 유형을 나타낸다[11-14]. 프로필 기반인 페이스북은 특정 사용자나 분야의 제한 없이 누구나 참여 가능한 서비스이다. 프로필 기반은 자신의 정보를 바탕으로 한 페이지가 다수 포함되어 있으며, 이를 타인과 공유할 수 있다.

연구 범위인 페이스북에서는 사용자 사이에서 서로 사회적 관계를 맺기 위해서는 '친구신청'을 했을 시 상대방이 수락해야 한다. 두 단계를 거치면 친구관계가

형성된다. 이런 과정은 양방향이다. 먼저 2.2에서 페이스북에서의 위협을 도출하고 연구 제안 방법을 후에 설명한다.

2.2 SNS에서의 악성 행위 및 개인정보 노출 현황

악성 행위를 사회심리학적으로는 플레이밍이라 말한다. 인터넷 기술의 발달로 인해 인터넷 플레이밍이 등장하게 되었으며, 이는 컴퓨터 환경에서 특정 사용자가 무례함이나 격렬함, 기쁨 등의 감정적인 행동을 분출하는 것으로, 적대감과 같은 강한 감정을 분출하는 경향에 이르기까지 부정적이면서 반사회적 행동을 포함하는 개념이다. 즉, 인터넷에서 이루어지는 악성 행위는 일상적인 오프라인에서 암묵적으로 합의되는 사회규범 등의 관행이 무시되는 경향이 있기 때문에 온라인에서 자연스레 이루어지는 것이다. SNS에서는 익명의 환경에서 자신의 신체와 신분이 노출되지 않아 발각의 위협이 낮으므로 공식적인 처벌에 대한 두려움이 작용하지 않으므로 여러 가지 악성 행위가 일어나고 있다. 또한 악성 행위는 익명성을 보장하거나 익명성을 보장하지 않는 것에 따라 빈도가 다른 것을 알 수 있다[15].

페이스북을 비롯한 대부분의 SNS에서는 특별한 인증 방법 없이 쉽게 가입이 가능하기 때문에, 익명성이 보장되어 다양한 악성 행위가 빈번히 일어나고 있다. Table 2.는 SNS에서 일어나고 있는 개인이 일으킬 수 있는 악성 행위의 종류를 나타낸다. 이는 [12,13]에 근거하여 재구성되었다.

악성 행위는 사용자의 개인정보가 노출되기 때문에 가능하다. 최근 세계 각국에서 SNS의 급속한 확산과 함께 개인정보보호에 대한 중요도가 증가하고 있다.

Table 1. The Function Types of SNS

구분	내용	서비스
프로필 기반	개인의 프로필 페이지를 기반으로 특정 사용자나 분야의 제한 없이 누구나 이용 가능한 서비스	페이스북, 싸이월드
비즈니스 기반	업무나 사업관계를 목적으로 하는 전문적인 비즈니스 중심의 서비스	링크드인, 비즈니스 페이스
블로그 기반	개인 미디어인 블로그를 중심으로 소셜 네트워크 기능이 결합된 서비스	이글루스, 네이버 블로그, 티스토리, 다음 블로그
버티컬 기반	사진, 동영상, 리뷰 등 특정 분야의 버티컬 UCC 중심의 서비스	유튜브, 다음팟
협업기반	공동으로 콘텐츠를 창작하는 서비스	위키피디아
관심주제 기반	분야별로 관심 주제에 따라 특화된 네트워크 서비스	도그스터, 트렌드밀
마이크로 블로그 기반	휴대전화 등을 이용하여 간단한 콘텐츠를 배포하는 서비스	트위터, 미투데이

Table 2. The Malicious Acts on SNS

종류	내용
악성코드	다양한 방법으로 악성 스크립트를 유포
광고성 글	스팸과 같은 광고성 글을 주기적으로 게재
단축 URL 피싱	단축 URL을 이용하여 의도적으로 피싱 사이트로 접근하게 만들
허위 사실 유포	정확하지 않은 정보를 유포시켜 사용자들 에게 혼란을 줌
사용자 위조	특별한 인증방법이 없는 것을 악용하여 사칭 계정 생성
악성 댓글	특정 사용자에게 욕설이나 비방글을 악의 적으로 게재
개인정보 유출	특정 사용자의 사생활 정보를 획득하여 유포
사용자 정보 수집	친구로 맺어진 사용자의 프로필 정보를 의도적으로 획득

SNS에서 이용자 개인의 정보가 유출되거나 악용될 가능성에 대한 우려가 높아짐에 따라, 2007년 유럽의 네트워크 및 정보보호기구인 ENISA는 이를 지적하면서, SNS 상의 가입자 프로필을 다른 제 3자가 데이터화할 목적으로 사적으로 저장하여 악용될 가능성이 있음을 지적했다.

Table 3.은 국내의 주요 SNS의 노출 가능 개인정보 현황을 나타낸다[16]. SNS에서는 다양한 개인정보나 중요 정보를 열람이 가능하며, 이런 정보를 조합하여 개인에 대한 기밀 정보를 쉽게 획득할 수 있다. 즉, 악의적인 사용자는 공개된 정보를 조합하여 개인 정보를 획득하여 악성 행위를 일으킬 수 있다.

Table 3. Exposed Personal Information on SNS

SNS 종류	공개된 신상정보 및 접근 가능한 정보 수준	
국내	미투데이	사진, 학교, 직장, 관심사, 지인 현황
	블로그	다음 블로그 : 사진, 자기소개 네이버 블로그 : 이름, 성별, 생일, 별명, 지역, 취미
	싸이월드	이름, 성별, 생년월일, 이메일, 혈액형, 집전화, 휴대전화, 집주소
국외	페이스북	이름, 성별, 생년월일, 이메일, 혈액형, 집전화, 휴대전화, 집주소
	트위터	이름, 거주지, 작성된 소개 정보

2.3 평판제도

본 논문에서 제안하는 신뢰도 평가 기준은 기존의 평판제도와 유사하다. 평판제도는 전자상거래가 발달

하면서 더 활발히 사용되고 있다. 각 서비스 업체들은 자신들의 서비스에 맞는 평판제도를 운영하고 있다. 온라인 평판제도는 기본적으로 사람들을 통해 얻는 정보에 대해 신뢰라는 요소를 포함한다. 평판 제도의 특성은 다음과 같이 설명할 수 있다[17,18].

- 평판제도의 행위자는 미래의 평판점수에 영향을 줄 수 있을 정도로 오랜 시간 활동해야 함.
- 현재의 점수체계는 미래의 점수에 영향을 줄 수 있도록 설계되어야 함.
- 과거의 행위에 대한 평가는 현재의 행위에 영향을 줘야 함.

평판제도는 보통 사기나 안전과 관련된 행위들을 구분할 수 있게 해준다. 평판을 조사하기 위한 검증 알고리즘은 온라인상에서 스팸이나 보안상의 문제, 악성 행위 등을 발견하거나 안전성에 대한 평가를 할 수 있도록 도와준다. 예를 들면, 인터넷에서 물건을 구매한 경우, 구매의사가 있는 사람이 이미 구매한 사람의 평가를 신뢰하고 구매 욕구를 충족시키는 것으로 드러났다. 즉, 신뢰가 없는 경우 사기 행위가 발생하기 쉬우며, 사기 행위자는 평판이 좋은 사용자의 신뢰평가를 가로채기도 한다. 이는 2.2에서 언급한바와 같이 인터넷 익명성에 따른 부정적 효과와 비슷하다.

사회심리학적으로 사람들은 자신이 속한 소셜 네트워크에서의 서로간의 영향력을 상호작용 하고 있으며 이는 개인의 활동을 더 증진시키고 어떠한 의사결정을 하거나 특정 사건에 대한 동기부여를 가져온다. 이러한 특성은 집단성과 관련하며, 사람은 자신이 속한 집단에 대해 신뢰를 가지며, 영향력을 서로 주고받는다. 평판제도에서 신뢰는 기본 요소이며, 온라인상에서 안정적인 접속과 접근성, 상호작용성, 그리고 개인화된 서비스와 커뮤니케이션이 신뢰에 긍정적인 역할을 미친다[19-21].

2.4 신뢰의 정의와 SNS의 사용자 신뢰관계

신뢰는 개인과 신뢰대상 간의 지속적인 관계에서 형성되는 주관적 믿음으로 정의된다. 신뢰는 사회적으로는 사회관계의 안정성을 유지하며, 사회체계의 잠재력을 확대하는 데 큰 역할을 수행한다. 온라인상에서의 신뢰는 경제적 교환에 대한 개념으로 발달되어왔다. 온라인 기반 SNS에서 의미하는 신뢰는 표현하려는 것이 정확하고, 옳은 것과 나쁜 것에 대해 구분이

확실하고 간단해야 하며, 믿음과 약속이라는 두 개의 요소를 지닌다. 즉, 서로가 친하다는 의미는 신뢰를 바탕으로 상호간의 관계가 성립하는 것이다. 또한 신뢰와 평판은 사기나 안전과 관련된 행위들을 구분할 수 있게 해준다. 이는 익명성이 전제되는 인터넷 환경에서 신뢰의 중요성이 점차 강조되었기 때문이다. 온라인에서의 신뢰 형성 과정은 작은 개인 정보에 기초하여 점차 커지면서 서로 간의 신뢰가 형성된다 [21-22].

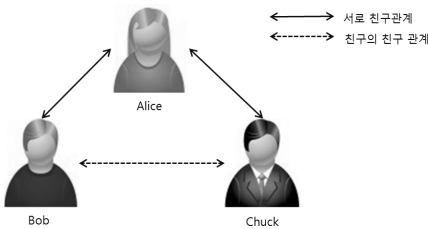


Fig.3. The Relationship

Fig.3.은 SNS에서의 신뢰관계를 설명한다. Alice, Bob, Chuck(이하 A, B, C) 세 사람이 있다. A와 B는 서로 친구이며 이는 신뢰관계가 형성됨을 뜻한다. 또한, A와 C도 친구관계이다. 하지만 B와 C는 친구관계가 아니지만, A와 B가 신뢰관계이기 때문에, B는 C에 대한 정보가 없어도 A로부터 C에 대한 정보를 획득하면, 그 말을 믿고 신뢰하게 되고 둘 사이에는 약한 신뢰관계가 성립된다. 마찬가지로, 같은 상황에서 FOAF(Friend of a friend, 친구의 친구)에 의해 Bob은 Chuck이 자신이 신뢰하는 Alice의 친구이기 때문에 Chuck과도 신뢰관계가 형성되는 것을 알 수 있다[23,24].

본 연구에서 의미하는 신뢰도 [21-24]와 같이 전자상거래에서 확장된 신뢰개념과 서로 상호작용을 하여 서로에게 영향력을 주는 것을 의미하고 사회통상적인 신뢰의 의미, 사기나 안전과 관련한 행위를 구분할 수 있는 요소로서의 의미도 포함한다.

III. 관련 연구

3.1 동적 사용자 신뢰도 평가 스킴

관련 연구 [26]에서는 본 논문에서 제안하는 바와 마찬가지로 사용자 계정에 신뢰도를 부여하는 방안을 제안하였다. SNS 상에서의 신뢰도를 평가하는 스킴

으로 Trust Parameter를 만들어 사용자의 신뢰도를 평가할 수 있는 동적 사용자 신뢰도 평가 스킴을 제안하였는데 Trust Parameter는 소셜 네트워크 서비스의 기능에 따라서 규범준수정도, 가치유사정도, 그리고 정보교류정도에 따라 신뢰도를 부여하는 방법을 제안하였다. Table 4.는 제안된 신뢰도 평가 스킴을 나타내며 서비스 기능의 항목을 바탕으로 신뢰도를 평가하는 방법을 제안하였다. 각 항목을 바탕으로 소셜 네트워크 서비스에서 전체적으로 공개될 공개 신뢰도와 사용자간의 신뢰도를 의미하는 관계 신뢰도 두 가지로 신뢰도 점수를 나타낸다.

하지만 [26]에서는 특정 SNS가 아닌 전체적인 관점에서 신뢰도 평가 스킴을 제안하였기 때문에 어떤 SNS에 적용할 수 있는지 언급하지 않았다. 따라서 기존의 연구를 좀 더 구체화하여 본 연구에서는 페이스북에 적용할 수 있는 평가기준을 제안한다.

Table 4. The Function Mapping Table

Trust Parameter	소셜 네트워크 서비스의 기능
규범준수정도	관계차단, 신고, 욕설 필터링
가치유사정도	추천, 스크랩, 동일그룹
정보교류정도	방문 횟수, 댓글, 메시지, 쪽지, 채팅

3.2 사용자 사이의 신뢰도 측정

SNS에서 사용자들 사이에 대해 신뢰도를 측정하는 연구는 많이 진행되었다. [5]에서 제안한 방법은 [24]의 연구처럼 SNS 상에서 FOAF 경우에는 신뢰관계가 성립한다고 정의하며, 이 경우에 친구의 친구가 실제 믿을 수 있는 사람인지를 판단한다. Alice, Bob, Chuck(이하 A, B, C) 세 사람이 있다고 가정할 때, A는 두 사람과 모두 친구이고 B와 C는 서로 친구관계가 아니다. 이 상황에서 B가 C에게 친구 요청을 했을 때, B가 어떤 사람인지 판단하기 위해서 A와 B 사이의 신뢰도를 측정하여 C에게 제공한다. 이 제안 방법의 주된 목적은 앞서 언급한 바와 같이, FOAF에게 친구 요청이 왔을 경우, 친구 요청자가 신뢰할 수 있는 사람인지 판단하기 위해서이다. 즉, 임의로 친구요청을 한 뒤, 관계가 성립되었을 때 사용자의 개인정보를 수집하는 악의적인 사용자를 막기 위해서 두 사용자 사이의 신뢰관계를 추론한다. 신뢰관계를 추론하기 위해 세 가지의 변수인 대화의 양, 대화의 빈도, 그리고 대화의 균형을 측정한다. 임계치를 설정하여, 측정된 신뢰도가 임계치에 도달하거나 넘으

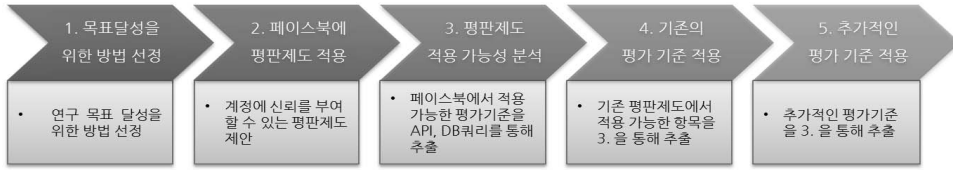


Fig.4. The Proposal Method of Trust Evaluation Criteria

면 신뢰관계인 사람으로 판정할 수 있다. 하지만 [5]에서는 FOAF가 아닐 때, 두 사용자 사이 관계를 측정하는 알고리즘을 제안하고 있지 않다. 따라서 연결고리가 없는 소셜 네트워크에 존재하는 사용자로부터 친구요청이 왔을 경우, 그를 판단할 수 있는 추가적인 기준이 마련되어있지 않기 때문에 모든 소셜 네트워크에서 사용자를 판단하는 것에 제한적이다.

IV. 페이스북에서의 사용자 신뢰도 평가 기준

4.1 신뢰도 평가 기준 제안 방법

본 연구의 목표는 페이스북에서의 안전한 정보 공유, 사용자간의 신뢰관계 보장, 그리고 악성 계정 판단을 위하여 기존에 운영 중인 평판제도를 이용하여

페이스북에서 구현 가능하여 적용할 수 있는 신뢰도 평가 기준을 제안한다. Fig.4.는 본 연구의 프로세스를 나타낸다. 먼저 페이스북의 API와 DB 쿼리를 분석하여 온라인 신뢰와 관련된 신뢰도 평가 항목을 추려내고, 기존의 연구에서 페이스북에 구현 가능한 항목을 추출한 뒤, 추가적인 평가기준을 제안한다. 따라서 본 연구에서는 온라인 신뢰와 연관된 페이스북에서 구현 가능한 항목으로 신뢰도 평가 기준을 제안한다.

4.2 페이스북 신뢰도 평가 기준 도출

페이스북에서 신뢰도 평가 기준을 도출하기 위해 페이스북에서 제공하는 오픈 플랫폼에서 그래프 API와 페이스북 쿼리 언어를 분석하였다. Fig.5.는 이를 분석한 매핑도이다. 신뢰도 평가에 적용할 수 있는 사



Fig.5. The Mapping between FB Open API and DB Queries

Table 5. The Evaluation Criteria list that can be applied on Facebook

항목	내용	적용 여부
앨범	사용자의 사진이나 사용자 친구의 사진이 모아진 페이지 <ul style="list-style-type: none"> • 앨범 ID, 만든 사람 ID, 앨범 타이틀, 앨범 설명, 앨범의 위치, 페이스북 앨범 링크, 앨범 커버 ID, 앨범의 개인 설정, 앨범의 수, 앨범의 타입(프로필, 모바일), 생성 시간, 최근 업로드시간 	○
코멘트	그래프 API로 구현 가능한 코멘트 창 <ul style="list-style-type: none"> • 코멘트의 페이스북 ID, 코멘트가 생성된 곳, 텍스트, 생성된 시간, 좋아요의 수, 댓글, 댓글의 수 	부분 적용
이벤트	이벤트에 대한 위치, 이름, 초대받은 사람, 참석자를 나타내는 항목으로 사용자와 페이지 객체는 이벤트 연결성을 가짐 <ul style="list-style-type: none"> • 이벤트 ID, 이벤트 생성자, 이벤트 이름, 이벤트 설명, 이벤트 시간, 이벤트 위치, 최근 업데이트, 이벤트 공개여부, 사진포함 	부분 적용
그룹	페이스북에서 그룹을 의미하며 사용자, 페이지 그리고 응용프로그램 객체가 하나의 그룹으로 연결됨 <ul style="list-style-type: none"> • 그룹의 ID, 그룹원, 그룹의 아이콘(친구의 그룹, 나의그룹), 그룹의 커버, 그룹의 소유자, 그룹의 이름, 그룹의 설명, 그룹의 웹사이트 링크, 그룹의 설정, 최근 업데이트 	부분 적용
친구 목록	페이스북에서 사용자의 친구 리스트로 그룹의 멤버 리스트와는 별개의 항목 <ul style="list-style-type: none"> • 친구 목록, 기타 공개여부는 권한 설정 가능 	○
포스트	프로필 피드를 대표하는 영역 <ul style="list-style-type: none"> • 포스트 ID, 포스트 게재한 사람, 포스트를 올린 곳, 메시지, 메시지 태그, 링크가 포함된 사진, 포스트에 포함된 링크, 링크의 이름, 링크의 설명, 플래시나 동영상, 액션(코멘트, 좋아요), 개인설정(접근제어), 포스트 위치, 텍스트, 텍스트 태그, 생성시간, 최근 업데이트 시간, 공유, 포스트의 타입(모바일, 웹 등) 	부분 적용
사용자 정보	사용자 프로필에 입력되는 모든 정보를 의미 <ul style="list-style-type: none"> • 사용자의 ID, 사용자의 이름, 성별, 장소, 나이, 업데이트 시간, 사용자의 계정 상태, 생일, 커버사진, 최근 설정, 로그인 장치, 학력, 이메일, 고향, 관심사, 현재 거주 위치, 결제 수단, 사진, 인용 문구, 종교, 보안 설정, 직장, 비디오 업데이트, 개인 웹사이트 주소 	부분 적용

용자 정보를 포함하는 항목은 위 Table 5.와 같다. 이는 페이스북에서 사용자의 신뢰도 평가 기준에 적용할 수 있는 부분을 도출한 것이다. 항목 간의 적용여부에 대한 내용을 살펴보면 먼저 앨범은 신뢰도와 연관 지었을 때, 다수의 사진을 업로드 한 이력은 자신의 프로필 사진이나 친구가 태그된 사진 등과 연관이 있고, 이러한 활동이 없는 사람보다 신뢰도를 더 가질 수 있다. 코멘트와 포스트에서의 신뢰도를 도출할 수 있는 부분은 페이스북에서 '좋아요' 버튼을 꾀 수 있다. 실제 유명인과 같은 허브 역할을 하는 사람일수록 다른 사람들이 그들의 포스트에 더 많이 관심을 가지는 것을 알 수 있으므로 '좋아요'의 수는 신뢰도를 평가하는데 좋은 기준이 될 수 있다. 이벤트의 경우에는 현실 세계에서의 모임을 초대하는 내용을 담고 있으므로 초대받은 사람은 실제 지인일 가능성이 높다. 즉 친구 관계라는 것은 신뢰를 기반 하므로 이벤트에 초대 받은 사람이나 참석자는 신뢰도를 부여할 수 있다. 또한 친구 목록은 서로의 친구 연결 고리를 나타내주는 항목이므로 신뢰도 평가 기준에 포함 가능하다. 그룹의 구성원들은 친구관계이거나 FOAF의 형태이고, 동일한 관심사에 대해 연결되어있으므로 신뢰도 평가

기준에 포함할 수 있다. 마지막으로 사용자정보는 개인이 자신의 프로필 사진이나 정보를 기입했을 때 추가적으로 신뢰도를 부여받는 것을 의미하는데 아무런 활동이나 자신의 정보를 기입하지 않는 사용자보다 사진과 자신의 정보를 기입해놓은 사용자가 더 높은 신뢰도를 받을 수 있다.

4.3 기존의 신뢰도 평가 기준 도출

본 절에서는 관련 연구인 페이스북 평판제도 제안 논문[25]과 평판제도를 도입한 SNS인 플러크[26]의 신뢰도 평가 기준을 도출한다. 평가 기준을 도출하는 방법은 Fig.5.를 통하여 페이스북에서 적용 가능한 항목을 추출한다.

[25]는 사용자 계정에 대한 평판 제도를 도입하는 것이 아닌 사용자와 그의 친구에 대해 신뢰도를 평가하는 방법이며, 평판 기준 항목과 내용은 Table 6.과 같다.

Table 6.에서 그룹의 수는 소셜 네트워크에서의 활동이 활발하다는 것을 의미하고 그룹에는 다양한 회원이 존재하며 이는 유대관계와 신뢰관계로 이루어져

Table 6. The list can be Evaluation Criteria from Related works

항 목	내 용	적용 여부
그룹의 수	그룹에 포함된 수나 활동에 따라 신뢰도 증가	O
함께 아는 친구	함께 아는 친구가 많으면 두 사람은 신뢰도가 있는 것으로 판단	O
메시지	친구들과 주고받는 메시지 수에 따라 신뢰도 증가	X
포스트	친구들과 주고받는 포스트 수에 따라 신뢰도 증가	X
사용자 정보	사용자 프로필에 입력되는 모든 정보를 의미하며 정보 추가 시 신뢰도 증가	부분 적용
친구의 수	친구의 수가 많을수록 신뢰도 증가	O

있으므로 신뢰도 평가기준에 적용이 가능하다. 함께 아는 친구와 사용자 정보는 앞 절에서 언급한 바와 같으므로 적용 가능하다. 메시지와 포스트는 계정 자체에 대해 평가하는 것은 무리가 있으며 사용자와 사용자 사이의 신뢰도 측정도 가능하다. 친구는 배경지식에서 언급한 바와 같이 신뢰관계가 형성되는데 신뢰관계의 연결고리가 많으면 많을수록 신뢰도를 높이 평가 받을 수 있다.

플러크(27)는 트위터와 같은 마이크로 블로깅 SNS이다. 현재 플러크에서는 Karma라는 평판제도를 도입해서 사용하고 있다. Karma는 플러크 상에서의 사용자의 행위 및 활동에 관하여 점수가 증가되고 감소된다. Table 7.은 Karma의 신뢰도 항목 및 본 연구에서의 적용가능 여부를 나타낸다.

플러크의 항목 중 적용 가능한 항목으로는 친구초대 시에 친구가 초대받지 않아 페이스북에 가입을 하는 것으로 이는 현실 세계의 지인일 경우에 가능하기 때문에 신뢰 관계가 전제되어 있다. 포스팅이나 댓글은 사용자와 사용자사이에서는 가능하지만 계정에 대해서 점수를 부여하는 것은 무리가 있다. 스팸 메시지는 페이스북에서도 스팸 메시지에 대한 신고기능이 있으므로 이는 신뢰도를 평가하는 데 좋은 항목이 될 수 있다. 사용자 정보는 앞서 언급했던 부분과 같고, 친구 요청을 거절당하거나 삭제당하는 것은 신뢰도와 관련이 있으므로 적용이 가능하다. 휴면계정은 오랜 기간 활동하지 않는 것으로 활동량이 많은 스팸머와 거리가 멀기 때문에 신뢰도 기준과 관련하여 페이스북에 적용하는 것은 옳바르지 않다.

Table 7. The list can be Evaluation Criteria from Plurk

항 목	내 용	적용 여부
친구 초대	친구를 초대했을 때 점수 상승	O
포스팅	자신의 담벼락에 글 게재 시 점수 상승	X
댓글	다른 글이나 자신의 글에 댓글 달 경우 점수 상승	X
스팸 메시지	부적절한 글을 올렸을 시 신고를 통해 점수 하락	O
사용자 정보	사용자 프로필에 입력되는 모든 정보를 의미하며 정보 추가 시 점수 증가	부분 적용
친구 요청 거절	친구 요청을 보냈으나 거절당할 시 점수 하락	O
친구 삭제	친구 관계가 삭제된 경우 점수 하락	O
휴면 계정	오랜 기간 동안 로그인 하지 않은 경우 점수 하락	X

4.4 추가적인 평가 기준 도출

본 절에서는 4.3에서 추가할 수 있는 신뢰도 평가 기준 이외의 새롭게 추가할 수 있는 평가 기준을 제안한다. Table 8.은 새로 추가되는 평가기준 항목을 나타낸다.

페이스북의 가입방법은 휴대폰 번호, 이메일 주소를 ID로 사용할 수 있다. 자신을 인증하는 것은 인증 절차를 따르지 않는 방법보다 자신의 신분을 노출하는 것이기 때문에 더 신뢰성을 가진다(15). 즉, 가입 방법에서 휴대폰 번호로 가입할 경우 추가 신뢰도를 줄 수 있다. 또한 계정의 나이는 일반적으로 악성 스팸머나 봇일 경우에는 계정의 나이 어린다는 연구 결과를 근거로 한다. 악성 계정이 유지하는 계정의 나이는 약 80일로 나타났다. 따라서 이를 통해 80일이 경과한 계정에 대해 더 높은 신뢰도를 줄 수 있다(28-29). 신고 횟수나 차단 및 삭제 횟수는 여타 평판제도에서

Table 8. Added Evaluation Criteria

항 목	내 용
가입방법	인증절차의 유무에 따라 점수 획득
계정의 나이	계정의 나이가 많을수록 추가 점수 획득
신고 횟수	신고의 횟수에 따라 점수 차감
차단 및 삭제	차단 및 삭제 횟수에 따라 점수 차감
상호관계 점수	자신과 연결된 친구들의 점수와 자신의 점수의 평균을 표기하여 점수 현황 알려줌

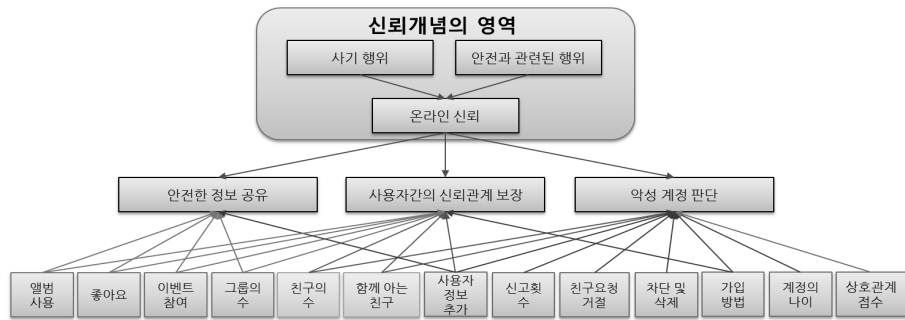


Fig.6. Under the concept of online trust, Research goals and Connectivity

많이 쓰이지만 페이스북에서도 신고제도의 기능이 있으므로 평가 기준에 추가할 수 있다[21][25-27]. 마지막으로 개인 노드의 평균은 자신과 연결된 친구들의 점수와 나의 점수에 대한 평균을 구하여 각자의 점수가 친구들에게 영향을 주게 함으로써 점수 관리에 더 의미 있는 활동을 하는 계기를 마련해준다. 이는 사회심리학에 근거한다. 사람은 상호작용을 하므로 서로의 영향력을 행사하는 것은 더 활발한 활동을 야기하는

것으로 알려져 있다[15-16][19-21]. 단, 점수는 개인의 점수와 자신과 연결된 노드의 점수 두 가지 모두 포기해야 할 것이다. 또한 자신의 점수에 친구의 점수가 영향을 주기 때문에 점수가 낮은 사람이 나에게 영향을 주게 되면 내 점수 역시 하락하기 때문에 점수가 낮은 사람을 찾아낼 수 있는 판단 기준을 마련할 수 있다. 결론적으로 점수가 낮은 악성 사용자일 경우에는 타인의 점수에 방해가 되기 때문에 프로필 기반

Table 9. The Proposed Trust Evaluation Criteria

구분	항목	내용
활동영역	앨범 사용	앨범에서는 개인의 사진이나 태그된 친구 등의 신뢰를 바탕으로 한 활동이 가능하므로 평가 기준에 적용 가능
	좋아요 수	좋아요 수가 많은 것은 영향력 있는 노드이며 영향력이 있다는 것은 신뢰할 수 있다고 판단할 수 있으므로 적용 가능
	이벤트 참여	이벤트는 현실 세계에서의 지인들이 모이는 자리를 의미하며 이는 신뢰관계를 바탕으로 하므로 적용 가능
	그룹의 수	그룹은 한 가지 이상 가입이 가능하고 그룹 자체는 동일하거나 비슷한 가치관을 가진 사람들이 서로 정보공유를 하므로 신뢰를 바탕으로 하므로 적용 가능
관계정도	친구의 수	악성 사용자는 친구의 수가 많지 않고 자신이 팔로워하는 수가 더 많기 때문에 친구의 수가 많은 것은 신뢰도가 높다고 판단할 수 있으므로 적용 가능
	함께 아는 친구	친구의 친구도 신뢰관계가 성립하므로 수치화하여 평가 기준에 적용 가능
계정관리	사용자 정보 추가	사용자의 정보가 추가되는 것은 그 계정의 활동성을 나타내므로 신뢰도 평가기준에 적용 가능
	신고 횟수	사용자가 악성 행위를 했을 경우에 일어나므로 행위를 기반으로 신뢰도 평가 기준에 적용 가능
	친구 요청거절	친구 신청을 거절당한 이유는 현실세계의 지인이 아닐 경우가 크기 때문에 행위를 기반으로 신뢰도 평가 기준에 적용 가능
	차단 및 삭제	악성 사용자일 경우에는 친구 노드가 삭제되거나 사용자가 차단될 수 있으므로 행위를 기반으로 신뢰도 평가 기준에 적용 가능
	가입 방법	페이스북에서 가입 방법은 휴대폰 인증과 이메일 인증이 있으며 휴대폰 인증은 사용자 자신을 드러내므로 이메일 인증 보다 더 높은 점수를 획득할 수 있고 신뢰도 평가 기준에 적용 가능
	계정의 나이	악성 사용자일수록 계정의 나이가 어리다는 것에 근거하여 계정의 나이가 일정기간 유지될 때 점수 획득이 가능하므로 신뢰도 평가 기준에 적용 가능
상호작용	상호관계 점수	자신과 연결된 친구들의 점수와 자신의 점수의 평균을 포기하여 점수 현황 알려주고 서로의 점수가 반영되므로 점수가 낮은 계정이 쉽게 드러나고 악성 사용자는 소셜 네트워크에서 고립 가능

SNS인 페이스북에서 친구가 모두 끊겨 어떠한 악성 행위를 해도 무의미해질 것으로 기대된다.

4.5 제안하는 신뢰도 평가 기준

페이스북에서 구현이 가능하여 적용할 수 있는 신뢰도 평가 기준과 기존의 연구, 그리고 추가한 평가 기준 항목을 정리하여 페이스북에서의 신뢰도 평가 기준을 제안한다. 신뢰도 평가기준은 Table 9.처럼 항목 간의 유사도에 따라 네 가지로 구분할 수 있으며 총 평가 기준 항목은 열 세 가지이다. 활동 영역은 사용자가 페이스북에서 활동하는 것을 바탕으로 측정할 수 있는 기준으로 앨범 사용, 좋아요의 수, 이벤트 참여, 그룹의 수를 포함한다. 관계 정도는 친구의 수와 함께 아는 친구의 수로 나누어진다. 친구는 서로 간의 관계를 의미하므로 신뢰관계로 맺어져있고 관련연구에서 언급한 바와 같이 FOAF도 신뢰관계가 형성되는 것에 근거하여 신뢰도 평가기준에 포함한다. 계정 관리는 사용자 계정에 대해서 신뢰도를 판단할 수 있는 것을 바탕으로 사용자 정보 추가, 신고 횟수, 친구 요청 거절, 차단 및 삭제, 가입 방법, 계정의 나이로 평가가 가능하다. 마지막으로 상호작용에서는 상호관계 점수 항목이 포함되어 있다. 이러한 항목들은 온라인 신뢰에서 요구하는 최소한의 정보를 통하여 나쁜 행위나 좋은 행위를 구분할 수 있다.

Fig.6.은 온라인 신뢰의 개념 하에 연구의 목표와 신뢰도 평가 기준의 연계성을 나타낸다. 신뢰의 핵심은 옳고 그름을 판단하는 것이다. 즉 악성행위나 안전과 관련된 행위를 판단할 수 있는 온라인 신뢰개념을 통해 연구 목표 세 가지를 달성할 수 있고, 이를 달성하기 위한 요소로 열 세 가지 신뢰도 평가기준을 제안

한다. 열 세 가지의 신뢰도 평가 기준은 각 목표와 연결이 되어있다. 활동 영역의 항목 전체는 첫 번째 목표인 안전한 정보공유에 포함된다. 이 항목은 서로가 연결된 노드끼리 정보 공유가 가능하기 때문이다. 또한 이 항목들은 신뢰관계를 보장하므로 전체가 두 번째 목표와도 연계성을 가진다. 관계정도 영역에서는 신뢰관계를 보장할 수 있고 이를 통해서 악성 계정 판단이 가능하므로 세 번째 목표와도 연계성을 띤다. 계정관리 영역에서는 전체 항목이 악성 계정을 판단할 수 있고, 사용자 정보추가는 사용자 자신을 인증하므로 앞의 두 가지 목표에도 포함된다. 가입 방법의 경우도 자신을 드러내기 때문에 사용자가 누구인지 판단이 가능하므로 신뢰관계가 보장된다. 마지막으로 상호작용 영역의 상호관계점수는 악성 계정을 쉽게 판단하기 위하여 제안되었으므로 세 번째 목표와 연계성을 가진다.

4.6 임의의 점수 산정

각 항목에 대한 점수를 평가하기 위해서 기호 정의가 필요하다. Table 10.은 구분에 따른 각 항목에 대한 기호 및 설명을 나타낸다. 또한 각 항목 간의 종속성을 고려하여 항목에 대한 가중치를 추가해야 한다. 표를 참고하여 제안 방법의 점수를 계산하는 방법은 (1)과 같다.

$$A = a(u)w_{a1} + l(u)w_{a2} + e(u)w_{a3} + g(u)w_{a4} \tag{1}$$

A는 활동 영역을 나타내며 A에 포함된 모든 항목의 값을 더하면 A를 도출할 수 있다.

Table 10. The Notations

구분	항목	기호	구분	항목	기호
활동영역 (A)	앨범 사용	$a(u)$	계정관리 (M)	사용자 정보 추가	$i(u)$
	좋아요 수	$l(u)$		신고 횟수	$n(u)$
	이벤트 참여	$e(u)$		친구 요청거절	$d(u)$
	그룹의 수	$g(u)$		차단 및 삭제	$b(u)$
관계정도 (R)	친구의 수	$f(u)$		가입 방법	$j(u)$
	함께 아는 친구	$foaf(u)$		계정의 나이	$a(u)$
상호작용 (I)	상호관계 점수	$r(u)$	기타	신뢰도	T
가중치 (w)	A영역 가중치	w_a		상호 신뢰도	T_i
	R영역 가중치	w_r		사용자 계정	u
	M영역 가중치	w_m			
	가중치 변수	i_n			

$$R = f(u)w_{r,1} + foaf(u)w_{r,2} \quad (2)$$

R은 관계정도를 나타내며 두 가지 항목을 더한 값을 의미한다.

$$M = i(u)w_{m,1} + n(u)w_{m,2} + d(u)w_{m,3} + b(u)w_{m,4} + j(u)w_{m,5} + a(u)w_{m,6} \quad (3)$$

M은 계정관리에 대한 항목들의 값을 의미한다. 따라서 세 가지 항목에 대한 사용자 u의 신뢰도 점수는 (4)와 같이 나타낼 수 있다.

$$T(u) = A(u) + R(u) + M(u) \quad (4)$$

마지막으로 상호작용에 대한 점수인 상호관계점수는 임의의 사용자 u와 연결된 모든 친구 노드들의 수인 n과 u의 평균 신뢰도 점수를 나타낸다. 즉, 사회심리학적 요소로 서로의 점수가 상호작용하는 것을 기반으로 한다. 이는 각자의 점수가 서로에게 영향을 미치므로 앞서 언급한 바와 같이 서로의 상호작용 활동량이 많아지는 것을 기대할 수 있다[19-21]. 식은 (5)와 같이 나타낼 수 있다.

$$I(u) = \frac{\sum_{i=1}^n T_i + T(u)}{n+1} \quad (5)$$

(5)인 상호관계점수의 제안 이유는 사람은 사회화 되어 있고, 사람들 사이에서 상호작용을 한다. 본 논문의 배경 지식, 2.3의 내용처럼 개인의 평판은 서로에게 영향을 주며, 이 영향력으로 사람들의 활동이 왕성해진다. 뿐만 아니라, 집단의 상호작용은 더욱이 활발한 활동을 기대할 수 있다. 따라서 나의 신뢰도점수가 친구에게도 반영되어 마지막 점수 산정 단계에서 포함되고 역으로 친구의 신뢰도점수가 나의 신뢰도 점수에 포함된다. 먼저 나의 신뢰도점수는 나와 내 친구들도 점수를 가지게 된다. 그 이후에 각자 가진 점수를 토대로 나와 연결된 사람들과 내 점수의 평균이 나에게 반영이 된다. 본 항목을 통해서 한 사용자는 자신과 친구 관계에 있는 사람 중 악성 행위를 하고 신뢰 점수가 낮은 사용자를 판단하고, 친구 관계를 끊을 수 있다. 제안한 페이스북의 신뢰기준에 대한 유효성을 판단하기 위해 [26]과 동일한 구성으로 실험을 진행하였다. 페이스북의 평균 소셜 네트워크 노드의 수인 130개의 노드를 임의로 구성하고 12개의 기준에 대한

Table 11. The Scope of Input Value

구분	항목	최소값	최대값
활동영역 (A)	앨범 사용	0	10
	좋아요 수	0	10
	이벤트 참여	0	10
	그룹의 수	0	10
관계정도 (R)	친구의 수	0	20
	함께 이는 친구	0	10
계정관리 (M)	사용자 정보 추가	0	20
	신고 횟수	-30	0
	친구 요청거절	-20	0
	차단 및 삭제	-10	0
	가입 방법	0	10
	계정의 나이	0	20
가중치 (w)	A영역 가중치	2	
	R영역 가중치	2	
	M영역 가중치	3	

점수를 랜덤 함수를 사용하여 각 노드들에 대한 점수를 측정한다. Table 11.은 각 기준별 수치에 대한 입력 범위를 나타낸다.

가중치는 각 항목마다 다른 수가 측정되어야 더 정확한 유효성을 가진다. 하지만 본 실험에서는 A, R, M에서의 가중치는 다르지만 항목에 대한 가중치는 임의로 같은 값으로 설정한다. 130개의 노드에 대해 랜덤함수를 이용하여 가중치와 수식을 통한 계산으로 나온 결과는 Fig.7.의 신뢰도 분포도와 같다.

최고 점수는 156점, 최하 점수는 -47점이었으며, 악성 계정은 그래프 상에서 명확히 드러나기 때문에 신뢰도 평가 기준으로 매겨진 점수를 통해 악성 계정

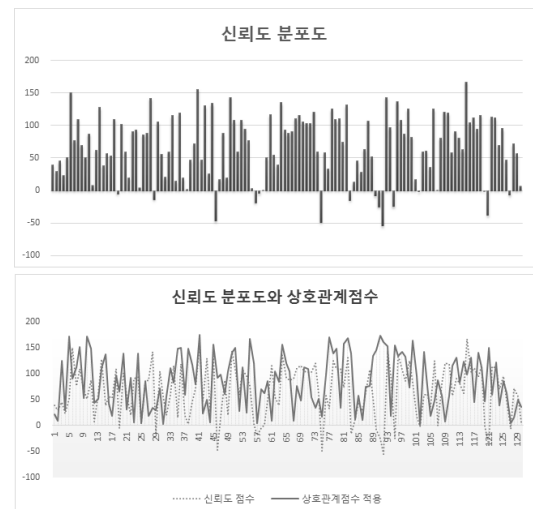


Fig.7. Trust Distributions and Interaction Score

을 쉽게 판단할 수 있다. 또한 아래의 그래프는 신뢰도 점수와 상호관계점수가 적용되었을 때, 점수변화를 나타낸다. 상호관계점수가 적용되면 한 노드의 점수가 올라갈 수도 있고, 내려갈 수도 있는 것을 알 수 있다. 따라서 악성 사용자의 행위로 인한 점수 변동으로 보다 쉽게 자신의 노드와 연결된 악성 사용자를 판단할 수 있다.

V. 기존 연구와 비교

기존 연구 [5]은 SNS 사용자와 사용자 사이의 대화의 양, 빈도, 그리고 균형에 따라서 신뢰도를 측정하여 FOAF인 경우에 신뢰점수를 알 수 있는 방법을 제안하였고, 이는 점수에 따라 신뢰구간과 비신뢰구간을 나눌 수 있다고 시각화된 그래프를 제시하였다. 그리고 페이스북의 평판제도를 제안한 논문인 [25]는 사용자 사이의 신뢰를 바탕으로 추출 가능한 항목을 평판제도를 제안하였다. [26]에서는 소셜 네트워크에서의 신뢰도 평가 스킴에 대해 제안하였다. 평가 항목은 세 가지의 Trust Parameter에 대해 서비스 기능에 따라 점수를 측정하는 방식이다. 기존의 SNS인 플러크[27]는 Karma라는 평판제도를 운영하고 있다. [5]에서 제안한 방법은 단순히 두 사용자 사이의 대화를 통해 신뢰도를 측정하므로 본 연구에서 제안하는 평판제도와는 관련성이 적고 제한적이다. 따라서 기존 연구인 [25], [26], [27]과 제안하는 기준을 페이스북에서 구현이 가능하고 적용 가능한지 알아본다. 페이스북 항목 이외에 새롭게 제안되어 추가된 신뢰도 평가 항목에 대해서도 매핑시킨다. Table 12.는 각

항목과 기존의 연구들, 제안하는 기준에 대한 매핑을 나타낸다. [26]에서는 규범준수정도, 가치유사정도, 그리고 정보교류정도에 대해 소셜 네트워크 서비스의 기능에 대해 신뢰도 평가 스킴 항목을 열린 가지 도출하였다. 본 연구와 [25]에서 제안한 항목 중 유사한 것은 세 가지이며, 기존의 연구는 페이스북에 적용하기에는 적합하지 않다. [26]에서는 페이스북에서 도출 가능한 항목 중 세 가지에 적합하고 추가 항목에서 한 가지 적용 가능하다. [27]에서는 페이스북의 항목 세 가지와 추가 항목에서 네 가지가 적용가능하다. 제안하는 기준은 총 열 세 가지가 페이스북에 적용 가능하므로 기존의 연구보다 더 정확하고 상세한 점수 측정이 가능할 것으로 기대된다.

평판제도를 효과적으로 운영하기 위해서는 이용자의 신뢰를 확보하고 네 가지의 보상을 제공해야한다. 금전적보상은 마일리지를 추가로 적립한다거나 추가적인 점수를 주는 것으로 이는 동기부여를 목적으로 한다. 심리적 보상은 점수가 낮은 사람, 즉 신뢰도가 낮은 사용자와의 상호작용은 떨어지고 반대의 경우는 상호작용이 높아지는 것을 근거로 한다. 명예적 보상은 참여도가 좋고 점수가 높은 우수사용자를 뜻한다. 마지막으로 조직·문화는 신뢰관계가 형성된 사용자들 간의 정보공유가 활발한 것을 의미한다[21]. 따라서 평판제도를 제안한 [25], [26], [27]과 본 연구에서 제안하는 기준의 효과를 입증하기 위해 평판제도에서 필요로 하는 네 가지 보상을 비교한다. 이는 관련 연구의 평가 기준 중에서 페이스북에 적용 가능한 항목만을 근거하여 제시한다.

Table 13.에서 [25]의 경우에는 사용자 사이의

Table 12. Comparison Applicability with Related works and Proposal method

		페이스북 항목 및 추가항목과 기존연구와 제안하는 방법의 적용 가능성 매핑			
기 존 페 이 스 북 항 목	항목	페이스북 평판제도 제안 [25]	동적신뢰도 평가스킴 [26]	플러크[27]	제안하는 신뢰도 평가 기준
		앨범			
	코멘트			○	○
	이벤트				○
	친구목록	○		○	○
	그룹	○	○		○
	포스트	○		○	○
추 가 항 목	사용자정보	○		○	○
	신고 횟수		○	○	○
	친구요청 거절			○	○
	차단및삭제		○	○	○
	가입 방법				○
	계정의나이				○
	상호관계점수				○

Table 13. Comparison of Acceptability Rewards

평판제도의 보상 수용 가능성				
보상 내용	[25]	[26]	[27]	제안 기준
금전적 보상			○	○
심리적 보상	○	○	○	○
명예적 보상			○	○
조직·문화	○	○		○

신뢰관계를 판단하므로 심리적인 보상과 조직 형성이 가능하다. [26]은 차단 및 삭제, 신고횟수를 적용할 수 있으므로 심리적 보상을 얻을 수 있고, 그룹 형성이 가능하므로 조직문화의 보상을 얻을 수 있다. [27]의 경우에는 활동량에 따라 점수를 부여하고 있으며, 스티커와 같은 것을 부여하고 우수 사용자 제도가 있으므로 금전적 보상, 심리적 보상, 그리고 명예적 보상을 받을 수 있는 것으로 나타났다. 마지막으로 제안하는 기준은 신뢰도 평가에 대한 점수를 산정하므로 점수라는 보상제도가 뒷받침하며 점수에 따라 사용자 간의 상호작용을 가능하게 제안하고 있다. 또한 점수를 통해 참여도가 높은 사용자를 판가름할 수 있고 신뢰관계 형성을 통해 서로간의 공유를 증진시킬 수 있으므로 조직·문화의 보상도 제공할 수 있다. 그러므로 평판제도의 효과를 입증하기 위한 필수 요소인 네 가지 보상을 제안하는 기준은 모두 수용할 수 있으므로 기존의 연구보다 더 객관적인 평판제도의 효과를 얻을 수 있다.

또한 2.3에서 언급한 바와 같이 여러 기업이나 사이트에서 현재 평판제도를 운영 중에 있으며, 여기에서 오는 긍정적 효과는 상호작용성을 기반으로 이루어지는 온라인 커뮤니케이션에서도 같은 효과를 얻을 수 있다. 이는 오프라인 환경 못지않게 사회적인 영향력을 발휘하기 때문이다. Fig.8.은 평판제도의 효과를 나타낸다[21].

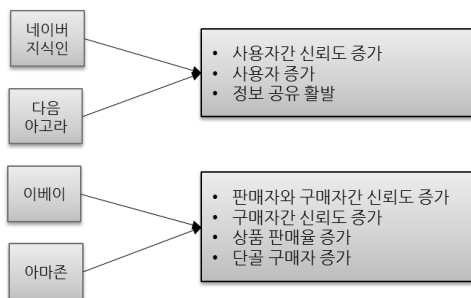


Fig.8. The Effect of Reputation Systems

즉, 하나의 작은 온라인 사회인 페이스북에서도 제한한 신뢰도 평가 기준을 통해 본 연구의 목표 세 가지를 포함한 긍정적 효과와 보상 제도를 내포하므로 신뢰성 있는 사용자 평가가 가능할 것으로 기대된다.

VI. 결 론

대체로 SNS는 개인정보 유출가능성이 높고 계정 생성 시 인증절차가 없기 때문에 많은 사람들이 배경의 제약 없이 쉽게 가입할 수 있다. 이러한 장점을 악용한 사용자와 악성 봇이 계정을 생성하여 불법 광고, 음란물 게시, 그리고 악성스팸트 삽입 등 다양한 악성 행위를 하고 있으며, 이는 빈번하게 일어나고 있다. 또한, 자유롭게 인맥형성을 하는 SNS에서 악의적인 사용자로부터 친구신청이 올 경우, 그 사용자를 판단할 수 없는 상황에도 많은 사람들이 친구 요청을 받아들여 개인정보가 쉽게 노출되고 있다. 이 경우에는 서로 친구관계인 양방향일 때, 개인정보가 유출될 가능성이 가장 높다. 이러한 SNS상의 문제점을 해소하기 위해 SNS의 사용자 사이에 신뢰도를 적용하는 연구와 다양한 SNS를 채택하여 개인정보가 추적되지 않게 하는 방법 등이 연구되었다. 하지만 기존의 연구의 단점을 보완하기 위해서 본 논문에서는 점유율이 가장 높은 SNS인 페이스북에서 안전한 정보 공유, 사용자 간의 신뢰관계 보장, 그리고 악성 계정 판단을 위해 온라인 신뢰개념을 채택하여 판단기준을 가질 수 있는 신뢰도 평가 기준을 제안하였다. 먼저 페이스북에서 구현 가능한 신뢰도 평가 기준 항목을 API와 DB 쿼리를 분석하여 추출하였고, 기존의 연구나 기존의 평판제도의 항목 중에서 페이스북에서 적용할 수 있는 항목을 추출하였다. 마지막으로 추가적인 평가 기준을 추출하여 열 세 가지의 페이스북에서 구현 가능한 신뢰도 평가기준을 제안하였다. 이를 이용하여 사용자 계정을 신뢰도 점수를 통해 판단할 수 있다. 또한, 신뢰도 점수가 친구들과 서로 상호작용하므로, 나의 점수를 낮게 만드는 악의적인 계정을 쉽게 걸러낼 수 있다. 이 방법이 계속 유지되면 결국 악성 계정은 소셜 네트워크에서 고립될 수 있다. 그렇기 때문에 SNS 상에서 일어나는 악성 행위가 최소화될 것으로 기대된다.

향후 연구에서는 제한한 신뢰도 평가 기준을 이용하여 각 항목에 대한 가중치를 비롯한 점수측정 방법을 항목 간의 상관관계를 염두에 두어 계산할 수 있는 방안을 추가 연구하여 실제데이터로 시뮬레이션 하고자 한다.

References

- [1] D. Boyd, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, 2007.
- [2] Mornitoring.com, "Facebook 2012 - Facts and Figures", <http://www.website-monitoring.com/blog/2012/10/19/facebook-2012-facts-and-figures-info-graphic>, Oct. 2012.
- [3] InternetTrend, "SNS Visitors Traffic," <http://www.internettrend.co.kr>, Oct. 2010.
- [4] Digital Journal, "Review: The latest Facebook scams to watch out for," <http://digitaljournal.com/article/337843>, Nov. 2012.
- [5] N. Chau, S.H Jung, "A two dimensional trust model based on communication behaviors of mutual friends in Social Networks," WISA, Aug. 2012.
- [6] American Statistical Association, "FACEBOOK REQUESTS FROM STRANGERS: Does Acceptance Differ by Gender of Requestor?," 2011.
- [7] B. Debatin, J.P Lovejoy, A.K Horn, and B.N Hughes, "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences," *Journal of Computer-Mediated Communication*, vol. 15, no. 1, pp. 83-108, 2009.
- [8] S. Leow, "You Don't Know Me But I Can Be Your Friend? Accepting Strangers as Friends in Facebook," pp. 34-40, 2009.
- [9] Pew Internet, "Why most Facebook users get more than they give," Feb. 2012.
- [10] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The Socialbot Network: When Bots Socialize for Fame and Money," ACSAC '11 Proceedings of the 27th Annual Computer Security Applications Conference, pp. 93-102, 2011.
- [11] Jinhjung. Lee, "SNS Diffusion and Trends," Korea Communications Agency, *Journal of Communication & Radio Spectrum*, vol. 44, pp.54-59, 2012.
- [12] Deokhoon Yu, Seungjoo Kim, "Research on Military SNS Protection for National Defense," *Journal of Korean Society for Internet Information*, 14(1), pp.41-52, Feb. 2013.
- [13] Symantec, "The Risks of Social Networking," Symantec Security Response, pp.1-28, 2010.
- [14] Soosang Lee, *Network Analysis Methods*, Nonhyeong, pp.370, 2013.
- [15] Kiyeong Noh, *Social media and cooperation in society*, Hanul Academy, pp.248-278, 2012.
- [16] The Korean Association for Public Society, *Social Network Services, Company and Politics*, pp.25-46, 2012.
- [17] J. Gibbs, and D. Giever, "Self-control and Its Manifestations Among University Students: an Empirical Test of Gottfredson and Hirschi's General Theory of Crime," *Justice Quarterly*, vol. 12, pp.231-235, 1995.
- [18] P. Resnick, and R. Zeckhauser, "Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system," Emerald Group Publishing Limited, pp. 127-157, 2002.
- [19] Myosub Um, *Personal and Society Interactions*, Novel Media, 2009.
- [20] Miyata Kakuko, *Society Relation Capital and the Internet Communication Books*, 2010.
- [21] National Information Society Agency, "Research on Reputation System for Online Trust," Feb. 2009.
- [22] C. Dwyer, SR. Hiltz, and K. Passerini, "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace," AMCIS, 2007.
- [23] FOAF, "FOAF Vocabulary Specification 0.98," <http://xmlns.com/foaf/0.1>
- [24] J.A Golbeck, "Computing and Applying

- Trust in Web-based Social Networks,” Ph.D. Thesis, Digital Repository at the University of Maryland, Apr. 2005.
- [25] W. Chen and S. Fong, “Social Network Collaborative Filtering Framework and Online Trust Factors: a Case Study on Facebook,” ICDIM, 2010 Fifth International Conference, pp. 266-273, 2010.
- [26] Changhoon Lee and Dongho Won, “Dynamic User Reliability Evaluation Scheme for Social Network Service,” Journal of The Korea Institute of Information Security & Cryptology, 23(3), pp.157-168, Apr. 2013
- [27] Plurk, “Plurk Karma,” <http://www.plurk.com>
- [28] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting Spammers on Twitter,” CEAS 2010 Jul. 2010.
- [29] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, “Detecting spam in a Twitter network,” First Monday, Jan. 2010

〈저자 소개〉



박 정 은 (Jeongeun Park) 학생회원
 2011년 2월: 대구가톨릭대학교 컴퓨터공학과 졸업
 2013년 8월: 고려대학교 정보보호대학원 정보보호학과 석사
 2013년 9월~현재: 고려대학교 정보보호대학원 정보보호학과 박사과정
 <관심분야> 정보보증, SNS 보안, 보안통계, 데이터마이닝, 사회공학



박 민 수 (Minsu Park) 학생회원
 2010년 2월: 신라대학교 컴퓨터네트워크학과 졸업
 2013년 2월: 고려대학교 정보보호대학원 정보보호학과 석사
 2013년 3월~현재: 고려대학교 정보보호대학원 정보보호학과 박사과정
 <관심분야> 정보보증, 정보보호제품 보안성 평가, 디지털 포렌식, Usable Security



김 승 주 (Seungjoo Kim) 종신회원
 1994년~1999년: 성균관대학교 정보공학과 (학사, 석사, 박사)
 1998년 12월~2004년 2월: KISA(舊한국정보보호진흥원) 팀장
 2002년~현재: 한국정보통신기술협회(TTA) IT 국제표준화전문가
 2004년 3월~2011년 2월: 성균관대학교 정보통신공학부 조교수, 부교수
 2011년 3월~현재: 고려대학교 사이버국방학과/정보보호대학원 정교수
 2004년~현재: 한국정보보호학회 이사
 2005년~2006년: 교육인적자원부 유해정보 차단 자문위원
 2007년: 국가정보원장 국가사이버안전업무 유공자 표창
 2007년~2009년: 전자 정부 서비스 보안 위원회 사이버 침해사고대응 실무위원회 위원
 2010년: 방송통신위원회 정보통신망 침해사고 민관합동조사단 위원
 2012년 3월~2012년 6월: 선관위 디도스 특별검사팀 자문위원
 <관심분야> 보안공학, 암호이론, 정보보증, 정보보호제품 보안성 평가, Usable Security