

개방형협업 참여자 기여도와 네트워크 특성과의 관계에 대한 연구 : 깃허브 오픈소스 프로젝트를 중심으로

Identifying the Network Characteristics of Contributors That Affect Performance in Open Collaboration : Focusing on the GitHub Open Source

백현미(Hyunmi Baek)*, 오세환(Sehwan Oh)**

초 록

발전된 정보통신기술은 개방형협업을 위한 오픈 플랫폼 역할을 수행함으로써 개인간 상호 협력을 용이하게 한다. 본 연구에서는 사회연결망이론을 바탕으로 개방형협업의 대표적 형태로 주목받고 있는 오픈소스 프로젝트의 상호협력 네트워크를 살펴보고자 한다. 특히 상호협력 네트워크 내에서의 개방형협업 기여도가 높은 참여자의 네트워크 중심구조를 분석해보고자 한다. 이를 위해 대표적인 오픈소스 프로젝트 플랫폼인 깃허브(GitHub)의 782개 리파지토리의 8,101명의 개발자를 대상으로 782개의 협업 네트워크를 도출함으로써, 협업에 있어 기여도가 높은 개발자의 연결·근접·매개 중심성 특성을 분석하였다. 분석 결과, 연결·매개·근접 중심성이 높은 개발자의 커밋수가 많아지는 것으로 나타났으며, 이중에서도 매개중심성에 의해 그 관계를 가장 잘 설명할 수 있는 것으로 나타났다. 또한 네트워크 사이즈가 클수록, 허브가 존재할수록 개발자의 매개중심성과 커밋수 간의 관련성이 더욱 커지는 것으로 나타났다. 본 연구는 향후 개방형협업의 성공적인 운영을 위한 시사점을 제시해줄 것으로 기대된다.

ABSTRACT

Information and communications technology facilitates collaboration among individuals by functioning as an open platform for open collaboration projects. In this regard, this study aims to understand the network characteristics of participants who contribute greatly to open collaboration by investigating the mutual cooperation network in an open source project, which represents a form of open collaboration based on social network theory. To achieve this objective, this study analyzes the network centrality of developers with a high number of commits, particularly 8,101 developers in 782 repositories in GitHub, a representative open source platform. This study also determines how the relationship between network centrality and number of commits depends on the size of a repository network and the presence of a hub. Consequently, the number of commits by developers with high degree,

이 논문은 2013년 한양대학교 교내연구비 지원으로 연구되었음(HY-2013-N).

* First Author, Department of Information Sociology, Hanyang University(lotus1225@hanyang.ac.kr)

** Corresponding Author, Korea International Trade Association(sehwano@kita.net)

Received: 2014-07-15, Review completed: 2015-02-09, Accepted: 2015-02-14

betweenness, and closeness centrality is increasing. Among which, betweenness centrality has the highest explanatory power. Furthermore, when a hub is present and as network size increases, the relationship between the betweenness centrality of a developer and his/her number of commits continues to grow. This study is expected to provide suggestions for the successful performance of open collaboration projects in the future.

키워드 : 개방형협업, 깃허브, 사회연결망분석, 개발자 기여도, 네트워크 중심성
Open Collaboration, GitHub, Social Network Analysis, Developer's Contribution, Network Centrality

1. 서 론

정보통신기술(ICT : Information Communication Technology)의 빠른 발전으로 인해 개인 간의 자발적인 조직 구성 및 프로젝트 진행이 용이해지고 있다. 발전된 정보통신기술은 개인의 창의적 아이디어를 생성·교류·공유·활용할 수 있는 오픈 플랫폼 역할을 수행하여 상호간 협력이 용이한 구조를 제공함으로써 개방형협업(Open Collaboration)을 가능하게 한다. 개방형협업이란 “내부 자원을 이용하던 방식을 벗어나서 인터넷을 이용하여 외부에 있는 다수의 자원을 이용하는 방식”으로 정의된다[19]. 과학자 집단과 주요기업을 연결함으로써 각종 연구개발 과제를 해결해줄도록 지원해주는 이노센티브(InnoCentive), 크라우드펀딩을 활용하여 개인이 원하는 비즈니스를 위한 펀드를 마련해주는 킥스타터닷컴(KickStarter.com), 오픈 백과사전인 위키피디아(Wikipedia) 등 다양한 형태의 개방형협업이 존재한다. 이처럼 집단지성을 통한 개방형협업의 활용 및 중요성이 증가하고 있음에도 불구하고, 개방형협업이 이루어지는 프로세스 및 네트워크 구조에 대한 연구는 비교적 제한된 편이다. 따라서 본 연구에서는 개방형

협업의 대표적인 형태인 오픈소스 프로젝트-다수의 개발자가 소프트웨어를 공동으로 개발함으로써 보다 나은 소프트웨어를 단기간에 개발할 수 있음을 목적으로 함-를 통해 성공적인 개방형협업을 위한 개발자간 네트워크 구조를 분석해보고자 한다. 특히, 본 논문에서는 오픈소스 프로젝트 개발자간의 협업에 있어 어떠한 네트워크 특성을 가진 참여자의 기여도가 높을지를 중점적으로 살펴보고자 한다.

Riehle et al.[29]은 오픈소스 프로젝트의 특징을 평등주의, 실력주의, 자발조직으로 규정하였다. 즉, 오픈소스 프로젝트는 누구든 접근이 가능하며 어떤 사람이든 기여할 수 있고(평등주의), 공헌자들은 그들의 실력을 바탕으로 투명하게 평가받을 수 있으며(실력주의), 프로젝트 커뮤니티는 어떻게 조직을 형성하고 프로젝트를 진행할지를 자발적으로 결정하게 된다는(자발조직) 특징을 가진다. 본 연구에서는 사회연결망이론을 바탕으로 성공적인 자발조직의 형태에 초점을 맞추어 공동연구 네트워크 형태와 참여자들의 자발적 참여에 대해 살펴보고자 한다. 이를 위해 오픈소스 프로젝트를 진행할 수 있는 대표적인 인터넷 기반 플랫폼 깃허브(GitHub)의 실제 데이터를 수집하여 실증분석을 실시하고자 한다. 2008년 만들어

진 깃허브는 2014년 340만 명 이상의 소프트웨어 개발자들에 의해 활용되고 있으며, 1,000만 개 이상의 리포지토리(repository: 프로젝트의 소스코드 저장소)가 운영되고 있다[40]. 깃허브는 개발자들이 자신의 활동에 관심 있는 사람에게 지속적으로 활동 내역을 알릴 수 있는 소셜 네트워크 사이트로서[37], 프로그래머들의 소셜 네트워크서비스라 불린다. 위키, 이슈, 코드 리뷰, 포크 등 다양한 기능을 제공함으로써 개발자에게 유용하게 평가받고 있으며 다수의 개발자들에 의해 활용되고 있다[12, 37].

연구의 도메인으로 깃허브를 선택한 이유는 대표적인 오픈소스 프로젝트를 수행할 수 있는 플랫폼이며, 또한 앞서 말한 바와 같이 깃허브에서 개발자들에게 제공되는 다양한 기능과 활동내역에 대한 정보가 API(Application Programming Interface, 응용 프로그램 프로그래밍 인터페이스)를 통해 제공되고 있기 때문이다. 즉, 깃허브 API를 통해 리포지토리의 커밋(commit: 리포지토리에 소스코드의 변경 내용을 반영하는 것)수, 개발자에 대한 정보 및 그들 간의 관계망 등 다양한 정보를 수집함으로써 개방형협업에 영향을 미치는 요인을 분석할 수 있다.

네트워크 특성과 성과에 대한 기존연구로는 기업 내에서의 커뮤니케이션, 조언, 정보의 흐름 등의 네트워크 구조와 성과간의 관련성에 대한 연구가 주로 많이 이루어졌다[7, 11, 20, 23, 35, 39]. 이에 더해 연구논문의 공동저자 네트워크 구조와 성과에 관한 연구[1, 2] 및 학생간의 커뮤니케이션, 우정, 공동학습 네트워크와 학생들의 성과(성적)에 관한 연구 등[6, 26, 28, 38]이 수행되었다. 본 연구에서 수행하고자 하는 오픈소스 프로젝트에

서의 개발자간 네트워크 구조와 성과에 관한 연구[15, 25, 34]도 수행된 바 있으나, 개방형협업의 중요성이 높아지는 현 시점에 비해 관련 연구는 부족한 실정이다. 특히, 기존 연구는 프로젝트를 분석 대상으로 하여 오픈소스 프로젝트의 성과와 네트워크 구조와의 관련성을 보고자 하는 연구가 주로 이루어졌으며, 개별적인 개발자의 성과에 초점을 맞춘 연구는 거의 이루어지지 않았다. 따라서 깃허브 오픈소스 프로젝트의 개발자간 상호협력 네트워크 구조를 통해 활동적인 개발자의 네트워크 특성을 살펴봄으로써 이를 보완하고자 한다.

이를 위해 깃허브의 782개 리포지토리에 속하는 8,101명의 개발자를 대상으로, 커밋수가 높은 개발자의 네트워크 중심성 특성을 분석하고자 한다. 탐색적 분석을 통해 개발자의 세 가지 네트워크 중심성(연결중심성, 매개중심성, 근접중심성)을 대상으로 개발자의 기여도와와의 관련성을 살펴본 후, 네트워크의 사이즈와 허브의 존재여부 등 리포지토리 특성의 조절효과를 살펴보고자 한다. 추가적으로 리포지토리의 운영자와 비운영자 간 네트워크 특성의 차이를 분석하고자 한다.

인터넷 플랫폼의 활용으로 인해 다양한 분야에서 등장하기 시작한 개방형협업은 향후 우리 사회의 주요한 조직형태로 자리잡을 것이다. 이러한 시점에서 자발적으로 형성된 협업 네트워크의 구조 및 중심적인 역할을 수행하는 참여자를 파악하는 것은 무엇보다 중요하다. 또한 앞서 말한 바와 같이, 기존 연구에서는 오픈소스 프로젝트의 성공적인 운영을 위한 각 개발자 성과에 초점을 맞춘 연구는 거의 다루고 있지 않다. 본 연구에서는 어떤 네

트위크 특성을 가진 개발자의 협업에 대한 기여도가 높은지를 파악함으로써 향후 성공적인 오픈소스 프로젝트 운영을 위한 경영, 정책, 플랫폼 디자인 등에 있어서의 시사점을 제공해 줄 수 있을 것으로 기대된다.

본 논문은 다음과 같이 구성된다. 다음 장에서는 본 연구의 이론적 배경과 관련 연구를 제시하고, 제 3장에서는 데이터 수집 및 기초 자료 분석에 대해 기술하고자 한다. 제 4장에서는 탐색적 분석 결과를 기술하며, 본 연구의 결과에 대한 토의와 함께 본 연구의 한계점과 향후 연구 등을 제 5장에서 다루고자 한다.

2. 기존연구 및 이론적 배경

2.1 기존연구

본 연구는 개방형협업을 위한 참여자간 공동 연구 연결망의 네트워크 특성과 개발자의 성과에 관해 살펴봄을 목적으로 한다. 최근 개방형협업이 활발히 이루어지고 있고, 그 중요성 또한 주목 받고 있지만 개방형협업에서 주요한 역할을 하는 자발적 참여자들의 네트워크 특성에 대해 살펴본 연구는 많지 않다.

네트워크 특성과 성과에 대한 연구는 다양한 분야에서 지속적으로 수행되고 있다. 대다수의 연구는 기업 내에서의 커뮤니케이션, 조연, 정보의 흐름 등의 네트워크 구조와 성과와의 관련성에 대한 연구이다[7, 11, 20, 23, 35, 39]. Cross and Cummings[7]는 석유화학회사의 엔지니어와 컨설팅회사의 컨설턴트를 대상으로 지식중심산업에서의 인식(awareness)과 정보(information) 네트워크 구조와 개인의 성과관

의 관계에 대해 살펴보았다. 분석 결과를 통해 두 개의 네트워크 모두에서 매개중심성이 높을수록 개인의 성과가 좋아짐을 확인하였다. Zhang and Venkatesh[39]는 직원들의 온라인과 오프라인상에서의 커뮤니케이션 네트워크 구조와 성과(job performance)에 대한 연구를 실시하였다. 통신회사 재직 104명에 대한 설문 조사를 바탕으로 분석한 결과, 온라인과 오프라인상에서의 커뮤니케이션 네트워크는 상호 보완적인 자원(complementary resources)으로서 성과에 영향을 미침을 밝혔다.

또한 학생간의 커뮤니케이션, 우정, 공동학습 네트워크와 학생들의 성과(성적)에 관한 연구도 다수 존재한다[6, 26, 28, 38]. Cho and Gay[6]는 컴퓨터 기반 협력학습(CSCL: computer-supported collaborative learning)에서 네트워크 특성과 학생들의 성과에 대한 실증 분석을 실시하였다. 분석 결과, 네트워크의 주요한 역할을 하는 학생들의 학업성과가 높게 나타남을 밝혔다. Ortiz et al.[26]은 네트워크 중심성(연결·매개·근접·아이젠벡터 중심성)과 학업성과에 대한 관계를 살펴보려고 하였다.

연구논문에 있어서의 공동저자 네트워크 구조와 성과에 관한 연구도 다수 존재한다[1, 2]. Abbasi et al.[1]은 정보시스템(information systems) 분야 학자들의 공동저자 네트워크(co-authorship network)를 살펴봄으로써 학자들의 연구역량과의 관련성을 분석하였다. Abbasi et al.[2]은 또한 철강 구조(Steel Structure) 분야 연구논문의 공동저자 네트워크를 통해 매개중심성이 높은 학자들을 중심으로 새로운 학자들과의 공동연구가 빈번히 일어남을 밝혔다. 경영정보 분야의 공동연구 패턴을 분석하기 위해 Huang et al.[14]은 MIS Quartely

의 공저자 분석을 통해 주요 연구 클러스트 및 주요 저자에 대한 분석을 시행한 바 있다.

본 연구에서 수행하고자 하는 오픈소스 프로젝트에서의 개발자간 네트워크 구조와 성과에 관한 연구도 수행되고 있으나, 대부분이 개발자의 성과보다는 프로젝트 자체의 성과에 초점을 맞추고 있다[15, 25, 34]. Kidane and Gloor[15]는 오픈소스 프로젝트에서 개발자의 밀도, 매개중심성과 같은 네트워크 구조와 그룹의 생산성간에는 상관관계가 존재함을 밝힌바 있다. Singh et al.[34]은 소스포지(SourceForge)의 프로젝트를 대상으로, 프로젝트 내 멤버 간 응집력(cohesion)이 강할수록, 외부 프로젝트와의 네트워크 응집력은 중간값(moderate)을 가질수록 프로젝트가 더욱 성공적임을 밝혔다. Oh and Jeon[25]의 연구에서는 오픈소스 커뮤니티의 실제 데이터와 시뮬레이션을 통해 오픈소스 커뮤니티 네트워크 사이즈와 연결성(connectivity)이 네트워크의 안정성에 어떤 영향을 미치는지를 살펴보았다.

2.2 사회연결망이론

최근 사회과학 분야에서 주목을 받고 있는 복잡계이론(CST : Complex System Theory)은 수많은 구성요소들이 상호작용을 통해 구성요소 하나하나의 특성과는 다른 새로운 현상과 질서가 나타나는 시스템을 의미한다. Simon [32]은 복잡계를 “단순하지 않은 방식으로 상호작용하는 많은 수의 부분으로 구성된 계”라고 정의하고 있다. 이러한 복잡계는 많은 수의 개체가 존재해야 하며, 이 개체들이 상호작용하여야 하며, 각각의 개체를 연구하는 것이 개

체들이 모인 계에 대한 이해에 큰 도움을 주지 못하는 경우를 의미한다. 따라서 이러한 복잡계이론은 개체 자체의 특성뿐만 아니라 네트워크 구조 내에서 상호작용하고 있는 개체의 특성을 살펴보는 것이 중요함을 시사한다. 이러한 복잡계로서의 연결망 사회를 살펴보기 위해, 본 연구에서는 관계적 인간관에 입각하여 인간 행위 및 사회 구조를 설명하고자 하는 사회연결망이론(SNT : Social Network Theory)을 이론적 배경으로 한다. 이 이론에 따르면, 각 개인의 상호작용 연결망은 행위를 통해 생산·유지되는 동시에 개인들에 의해 생겨나는 연결망 구조는 그들의 행위에 영향을 미친다[17]. 다시 말해, 본 이론에 따르면 인간은 상호 연결망에서 자신의 네트워크 위치 및 연결망의 구조에 따라 다르게 행동한다[17]. 사회연결망이론에서는 사회적 관계를 네트워크 안에서의 개별적인 개체를 의미하는 노드와 개체들 간의 관계를 의미하는 링크의 관점에서 바라보고자 한다[22]. 이러한 사회연결망적 접근은 개체간의 상호작용을 노드를 나타내는 그래프의 점과 링크를 나타내는 선으로 연결하여 설명한다[21, 24].

최근 들어 이러한 사회적 관계를 분석할 수 있는 사회연결망분석(SNT : Social Network Analysis)을 활용한 연구가 많이 등장하고 있다. 이는 연결망 내에서의 개체간 상호작용을 중심으로, 개체간 연결상태 및 연결구조의 특성을 계량적·시각적으로 표현하는 분석기법이다[16, 27]. 협업에서의 사회연결망분석은 참여자 개별적 속성뿐만 아니라 참여자 상호간 관계적 속성을 대상으로 분석이 이루어지므로 상호작용에 의한 협업의 기본적인 성격을 반영할 수 있는 적절한 방법이라 판단된다. 일반

적으로 사회연결망 분석은 네트워크 특성 분석, 중심구조 분석, 응집구조 분석, 역할구조 분석의 4가지 방법을 통한 접근이 가능하며, 본 연구에서는 이러한 분석 방법 중, 개별적인 노드 단위의 중심구조 분석을 실시하여 네트워크 구조 내에서 중요한 역할을 수행하는 노드를 파악하고자 한다. 네트워크 중심성으로는 크게 연결중심성(Degree Centrality), 매개중심성(Betweenness Centrality), 근접중심성(Closeness Centrality)의 세 가지를 들 수 있으며[30], 각각의 중심성이 의미하는 바는 다음과 같다.

연결중심성(degree centrality)은 한 노드에 직접적으로 연결되어 있는 노드의 개수를 의미하며, 연결된 노드의 수가 많을수록 연결중심성이 높아지게 된다. 특정 노드 k 의 표준화된 연결중심성은 다음의 수식과 같다[31].

$$C_D(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n-1}$$

본 수식에서 n 은 네트워크 내에 있는 노드의 수를 의미하며, $a(p_i, p_k)$ 는 노드 i 와 노드 k 가 연결된 경우는 1, 연결되지 않은 경우 0의 값을 가지게 된다. 표준화를 위해 이론적으로 가능한 최대값인 $(n-1)$ 로 나누어준다.

매개중심성(betweenness centrality)은 한 노드가 네트워크 내의 다른 노드들 사이의 최단 경로 위에 위치할수록 그 노드의 매개중심성이 높아지며, 특정 노드 k 의 표준화된 매개중심성을 수식으로 표현하면 다음과 같다[4].

$$C_B(p_k) = \frac{\sum_{i < j}^n \frac{g_{ij}(p_k)}{g_{ij}}}{(n-2)(n-1)}$$

본 수식에서 g_{ij} 는 노드 i 와 노드 j 사이의 가장 짧은 거리(geodesic distance)를 의미하며, $g_{ij}(p_k)$ 는 노드 k 를 포함한 상태에서 노드 i 와 노드 j 사이의 가장 짧은 거리를 의미한다.

근접중심성(closeness centrality)은 한 노드로부터 다른 노드에 도달하기 위해 필요한 최소 단계의 합의 역수로 정의되며, 근접중심성이 높을수록 네트워크 내의 모든 노드에 도달하기 위한 단계가 작아지게 된다. 특정 노드 k 의 표준화된 근접중심성은 다음의 수식으로 표현된다[10].

$$C_C(p_k) = (n-1) \sum_{i=1}^n d(p_i, p_k)^{-1}$$

본 수식에서 $d(p_i, p_k)$ 는 노드 i 와 노드 k 사이의 최단 경로를 의미한다. 즉, 모든 다른 노드들까지의 최소거리의 합이 최소가 되는 경우 근접중심성이 가장 높은 노드라 할 수 있다.

3. 데이터수집 및 기초자료분석

3.1 데이터수집

본 연구의 분석대상이 된 깃허브 리파지토리는 데이터를 수집한 2014년 3월 17일을 기준으로 만들어진 지 1년 가량 된 리파지토리(2013년 1월 1일부터 2013년 2월 28일 기간 동안 생성된 리파지토리) 중, 스타어링(Starring: 개발자가 관심이 있는 특정 리파지토리를 스타어링하면 그 리파지토리가 북마크됨)의 수가 5개 이상(특정 리파지토리를 즐겨찾기 한

개발자가 5인 이상인 리파지토리이다. 이러한 기준에 적합한 11,503개의 리파지토리를 대상으로 리파지토리 및 개발자 관련 정보를 수집하였다. 추가적으로 공동연구 네트워크를 형성하기 위해, 타겟 리파지토리의 각 이슈(issue)를 대상으로 하나의 이슈에 함께 토의에 참여한(코멘트를 남긴) 개발자들에 대한 정보를 수집하였다. 이를 위해 파이썬(Python)을 활용한 웹크롤러(web crawler)를 구축하여 깃허브 API 및 깃허브 웹페이지를 활용하여 정보를 수집하였다. 리파지토리에 소속된 각 이슈에 대해 토의를 한 개발자들을 대상으로 상호협력이 이루어졌다고 가정하고[8], 사회연결망분석을 위한 개발자-이슈간 2-모드 네트워크 데이터 형태를 구축하였다. 사회연결망분석에서 2-모드 네트워크란 서로 성질이 다른 개체 간의 관계로 구성된 네트워크를 의미하며, 본 연구에서는 이슈-개발자의 상이한 개체간 네트워크로 구성된다. 수집된 11,503개의 리파지토리가 모두 활성화되지는 않았으며, 실질적으로 공동연구가 아닌 개인연구 수준에서 리파지토리가 운영되는 경우도 많이 존재

한다[18]. 따라서 본 연구에서는 앞서 선택한 11,503개의 리파지토리 중 사회연결망 분석이 가능한 수준인 하나의 리파지토리 내에서 이슈 코멘트가 100개 이상인 782개의 리파지토리를 최종 분석 대상으로 하였다[8].

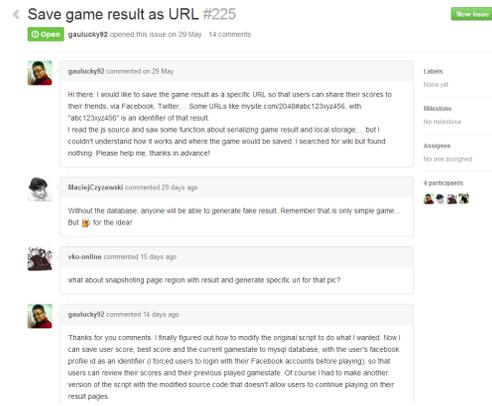
<Table 1>은 본 연구를 위해 깃허브에서 수집된 변수에 대해 설명하고 있다. 본 연구에서 사용된 종속변수인 기여도는 기존연구에 따르면, 공통적인 목적을 달성하기 위해 각각의 구성원들이 업무와 관련된 행동을 하는 것을 의미하며[33], 본 연구에서는 각 개발자의 특정 리파지토리에서의 커밋수로 이를 측정하고자 하였다. 즉, 본 연구에서는 개발자 i 가 리파지토리 j 에서 이룬 기술적 성과인 커밋수(Commit_{ij})를 개발자 i 의 리파지토리 j 에 대한 기여도라 정의하였다. 이러한 커밋수는 소스 코드의 의미 있는 변화를 반영하기 때문에 기존 연구에서도 개발의 성과(기술적 성과)를 위한 측정(measure)으로 이미 활용된 바 있다[13].

본 연구의 독립변수는 협업을 위한 목적으로 참여자간 관계망이 형성됨으로써 도출된 네트워크 특성으로, 개발자 수준의 네트워크 특

<Table 1> Data Collected from GitHub

Variable		Definition
Dependent Variable	Commit _{ij}	Number of developer i 's commits in repository j
Independent Variable	DegCentrality _{ij}	Degree centrality of developer i in repository j
	BetCentrality _{ij}	Betweenness centrality of developer i in repository j
	CloCentrality _{ij}	Closeness centrality of developer i in repository j
Control Variable	Period _i	Working period of developer i in GitHub (unit : day)
	RepoNumber _i	Number of repositories where developer i is participating
	TotalCommit _j	Total number of commits in repository j
	Size _j	Size of repository j (unit : kilobyte)

성을 측정하기 위해 각 개발자의 리파지토리 내에서의 세 가지 중심성 지수를 사용하였다. 앞서 정의한 연결중심성(DegCentrality_{ij}), 매개중심성(BetCentrality_{ij}), 근접중심성(CloCentrality_{ij}) 지수를 사용하였다. 본 연구의 독립변수로서 사용된 세 가지 중심성 지수는 리파지토리별 개발자-이슈간 2-모드 네트워크를 개발자간 1-모드 네트워크로 변환하여 네트워크를 구성한 후, 협업 네트워크 내에서의 개발자들의 네트워크 중심성을 도출하였다. 즉 최종적으로 분석대상으로 선정된 782개 리파지토리 각각의 개발자간 협업 네트워크를 구성함으로써(리파지토리별로 782회의 사회연결망 분석을 실시함으로써) 도출된 값들이다. 개발자가 갖는 특성 및 리파지토리가 가지는 특성을 통제하기 위해 통제변수로는 개발자의 활동기간(Period_i), 개발자의 활동 리파지토리수(RepoNumber_i) 및 리파지토리의 전체 커밋수(TotalCommit_j), 리파지토리의 사이즈(Size_j)를 고려하였다. <Figure 1>은 깃허브에서 리파지토리 내에서 하나의 이슈에 대해서 개발자간 코멘트를 공유한 화면을 보여준다.



<Figure 1> Snapshot of Issue Comments in a GitHub Repository

3.2 기초 자료 분석

분석대상인 리파지토리 782개에 소속된 개발자들 중 종속변수로 쓰인 커밋수(Commit_{ij})가 평균에서 3표준편차 이상 떨어지지 않은 개발자만을 대상으로 분석을 실시하였다. 이 중 관련 변수에 대한 데이터를 모두 확보할 수 있는 8,101개의 데이터(782개의 리파지토리에 소속된 8,101명의 개발자에 대한 데이터)

<Table 2> Descriptive Statistics

	N	Minimum	Maximum	Mean	Standard Deviation
Period _i	8101	0	2,340	1,195.31	557.05
RepoNumber _i	8101	0	839	41.22	62.92
TotalCommit _j	8101	1	10,000	947.88	1,618.01
Size _j	8101	51	1,518,764	36,670.29	124,194.1
DegCentrality _{ij}	8101	0	0.489	0.027	0.037
BetCentrality _{ij}	8101	0	1	0.124	0.244
CloCentrality _{ij}	8101	0	1	0.555	0.202
Commit _{ij}	8101	1	753	56.37	117.76

<Table 3> Results of Correlation Analysis

	1	2	3	4	5	6	7	8
1. Period _i	1.000							
2. RepoNumber _i	0.390**	1.000						
3. TotalCommit _j	-0.181**	-0.136**	1.000					
4. Size _j	-0.107**	-0.095**	0.359**	1.000				
5. DegCentrality _{ij}	0.070**	0.079**	-0.142**	-0.081**	1.000			
6. BetCentrality _{ij}	0.099**	0.106**	-0.132**	-0.078**	0.297**	1.000		
7. CloCentrality _{ij}	0.121**	0.131**	-0.104**	-0.097**	0.498**	0.732**	1.000	
8. Commit _{ij}	0.040**	0.017	0.111**	0.045**	0.179**	0.561**	0.505**	1.000

* : p < 0.05, ** : p < 0.01.

를 대상으로 분석을 시도한 결과(<Table 2> 참조), 개발자의 특정 리포지토리에서의 평균 커밋수(Commit_{ij})는 56개, 평균 연결중심성(DegCentrality_{ij})은 0.027, 매개중심성(BetCentrality_{ij})은 0.124, 근접중심성(CloCentrality_{ij})은 0.555의 값을 가짐을 확인할 수 있었다. 개발자들의 평균 깃허브 활동기간(Period_i)은 1,195일이었으며, 평균적으로 활동하는 리포지토리의 수(RepoNumber_i)는 41개였다. 분석에 사용된 리포지토리의 평균 전체 커밋수(TotalCommit_j)는 948개였으며, 사이즈(Size_j)는 36,670킬로바이트인 것으로 나타났다. <Table 2>에서는 본 연구에서 활용되는 기본적인 변수의 기술통계 분석 결과를 확인할 수 있으며, <Table 3>에서는 변수간 상관관계 분석 결과를 확인할 수 있다.

4. 분석 결과

본 장에서는 깃허브 데이터를 기반으로 탐색적 분석을 통해 개발자의 리포지토리 협업 네트워크에서의 중심성과 개발자의 리포지토리 내

에서의 기여도와의 관련성을 살펴보고자 한다. 또한 리포지토리의 특성인 네트워크 사이즈와 허브 존재여부가 앞서 도출한 관련성에 어떤 영향을 미치는지를 살펴보고자 한다. 추가적으로 리포지토리의 운영자와 비운영자 간 네트워크 중심성에 대한 차이를 분석하고자 한다.

4.1 개발자 네트워크 중심성과 기여도와의 관련성

각 리포지토리 공동연구 네트워크 내에서의 연결·매개·근접 세 가지 중심성과 개발자의 커밋수와의 관련성을 살펴보기 위해 각 모델에서의 표준화 계수값을 산정하여 비교해 보았다(<Table 4> 참조). 주목할만한 점은, 세 가지 중심성 관련 모델 모두에서 1차항 모형에 비해 2차항 모형에 의한 설명력이 모두 증가하는 것으로 나타났다. 이를 통해 세 가지 중심성 모두 개발자의 커밋수에 미치는 영향이 선형이 아닌 비선형임을 확인할 수 있다. 또한 세 가지 중심성 중, 매개중심성이 1차항, 2차항 모형 모두에서 가장 설명력이 큰 것으로 나타났다. 이는 논문의 공동저자 네트워크에서 매

<Table 4> Relationship between Network Centrality and Developer's Commit : First Order Equation versus Second Order Equation

First Order Equation	Standardized Coefficient		Adjusted R ²	
	DegCentrality _{ij}	0.17***		0.032
BetCentrality _{ij}	0.56***		0.314	
CloCentrality _{ij}	0.50***		0.254	
Second Order Equation	1 st order		2 nd order	Adjusted R ²
	DegCentrality _{ij}	0.36***	-0.22***	0.048
	BetCentrality _{ij}	1.04***	-0.51***	0.334
	CloCentrality _{ij}	-0.31***	0.86***	0.316

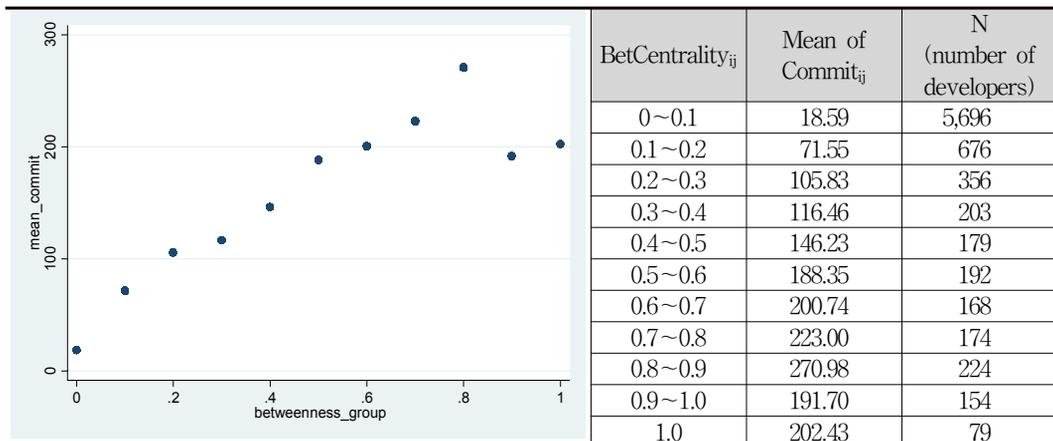
* Dependent variable : Commit_{ij}.
 * : p < 0.05, ** : p < 0.01, *** : p < 0.001.

개중심성이 높은 학자일수록 새로운 학자들과의 공동연구가 더욱 빈번하게 이루어진다는 Abbasi et al.[2]의 연구결과 및 기업에서의 정보 흐름 네트워크에서 매개중심성이 높은 직원의 성과가 높게 나타난다는 Cross and Cummings[7]의 기존 연구결과 등과 동일한 결과라 볼 수 있다. 즉, 본 분석 결과를 통해 오픈소스 프로젝트의 개방형협업 네트워크에서도 매개중심성이 높은 개발자들의 역할이 중요함을 확인할 수 있었다. Burt[5]에 따르면, 매개중심성이 높은 노드는 다른 노드들을 서

로 연결해주는 중개인(broker) 역할을 수행함으로써 네트워크 내에서의 정보 흐름을 제어하는 역할을 수행하게 된다. 이러한 이유로 다른 노드의 입장에서는 매개중심성이 높은 노드에 대한 의존성이 높아지게 되고, 결과적으로 매개중심성이 높은 노드는 네트워크 내에서의 영향력이 강해지게 된다.

분석 결과를 통해 두 개의 네트워크 모두에서 매개중심성이 높을수록 개인의 성과가 좋아짐을 확인하였다. 매개중심성의 경우 표준화계수값이 1차항의 경우 0.56으로 나타났

<Table 5> Betweenness Centrality and Developers' Average Number of Commits



으며, 2차항 모형에서는 2차항의 표준화 계수 값은 -0.51, 1차항의 표준화 계수값이 1.04로 나타났다. 2차항 모형의 형태를 자세히 살펴 보기 위해 <Table 5>는 매개중심성을 0.1 단 위로 나누어서 각 그룹에서의 개발자 커밋수의 평균값을 그래프로 나타내었다. 그래프에 따르면, 매개중심성과 개발자 커밋수간 관계는 1차항이 아닌 2차항을 따르고 있으며, 매개중심성이 0.8인 경우 개발자 커밋수가 가장 높음을 확인할 수 있다.

연결·매개·근접 세 가지 중심성의 경우 상호간의 상관계수가 높으므로(<Table 3> 참조), 하나의 모델에서 이를 고려할 경우 다중공선성 문제가 발생할 수 있다. 따라서 본 연구에서는 세 가지 중심성 중 가장 영향력이 높은 매개중심성을 독립변수로 한 모형을 최종적으로 만들고자 한다. 최종 모델에서는 리파지토리 내에서의 개발자커밋수(Commit_{ij})에 미치는 통제변수로서 개발자활동기간(Period_i), 개발자활동 리파지토리수(RepoNumber_i), 리파지토리

커밋수(TotalCommit_j), 리파지토리사이즈(Size_j)를 고려하였고, 독립변수로는 개발자의 리파지토리 공동연구망 내 매개중심성(BetCentrality_{ij})과 매개중심성의 이차항(BetCentrality_{ij}²)을 고려하였다. 분석을 위해서는 위계적 회귀분석 모형을 사용하였다. 위계적 회귀분석모형은 연구자의 연구목적에 맞게 독립변수를 추가적으로 투입해가며 그 영향력을 검증하는 모델이다. <Table 6>은 위계적 회귀분석 결과를 정리한 표이다. 모델 1에서는 리파지토리 커밋수를 제외한 세 가지 통제변수의 영향력을 살펴보았으며, 모델 2에서는 통제변수 중 개발자 커밋수에 가장 많은 영향을 미치는 리파지토리 커밋수를 추가로 고려하였다. 모델 3에서는 독립변수인 매개중심성을 추가로 투입하였으며, 모델 4에서는 최종적으로 매개중심성의 2차항까지 고려하였다. 모델 분석 결과, 개발자 활동기간이 길수록, 개발자가 활동하고 있는 리파지토리 개수가 작을수록, 리파지토리의 사이즈 및 전체 커밋수가 많을수록 각 개발자의 리파지토리

<Table 6> Output of Hierarchical Regression Model

	Model 1	Model 2	Model 3	Model 4
(Constant)	43.213***	32.178***	2.795	-0.930
Period _i	0.009***	0.012***	0.006**	0.004*
RepoNumber _i	0.008	0.021	-0.056**	-0.063***
Size _j	0.000047***	0.000008	0.000025**	0.000025**
TotalCommit _j		0.009***	0.013***	0.013***
BetCentrality _{ij}			282.674***	527.01***
BetCentrality _{ij} ²				-318.032***
R ²	0.004	0.016	0.351	0.372
Adjusted R ²	0.004	0.016	0.350	0.372
R ² change	0.004***	0.012***	0.334***	0.021***
N	8,101	8,101	8,101	8,101

* Dependent variable : Commit_{ij}.
 * : p < 0.05, ** : p < 0.01, *** : p < 0.001.

내에서의 커밋수가 증가하는 것으로 나타났다. 또한 매개중심성은 역U자(inverted-U)형태로 커밋수와 관련성이 있음을 확인할 수 있다 (<Table 5>, <Table 6> 참조).

4.2 조절효과 분석 결과 : 협업 네트워크 사이즈 및 허브 존재여부

협업 네트워크 사이즈 및 허브 존재여부에 따라 개발자의 매개중심성과 커밋수와의 관계가 어떻게 달라지는지를 살펴보기 위해, 리파지토리 네트워크 사이즈와 네트워크 내 허브 존재유무에 의한 그룹을 중심으로 조절효과를 분석하였다.

4.2.1 협업 네트워크 사이즈

분석대상인 리파지토리의 협업 네트워크 사이즈는 평균 0.428로 나타났으며, 이 평균 값을 기준으로 네트워크 사이즈가 큰 리파지토리와 작은 리파지토리로 나누어 분석을 실시하였다. 1차항 모형(위계적 회귀분석 모델 3)을 분석한 결과, 네트워크 사이즈가 큰 경우 매개중심성과 커밋수와의 관련성이 더욱 강해짐을 확인할 수 있다(<Table 7> 참조). <Table 7>은 1차항 모형을 대상으로 두 가지 그룹에서의 매개중심성과 개발자 커밋수 관계에 있어서의 차이점을 분석하였으며, <Table 8>에서는 2차항 모형(위계적 회귀분

<Table 7> Analysis Results of Moderating Effects : Network Size (First Order Equation)

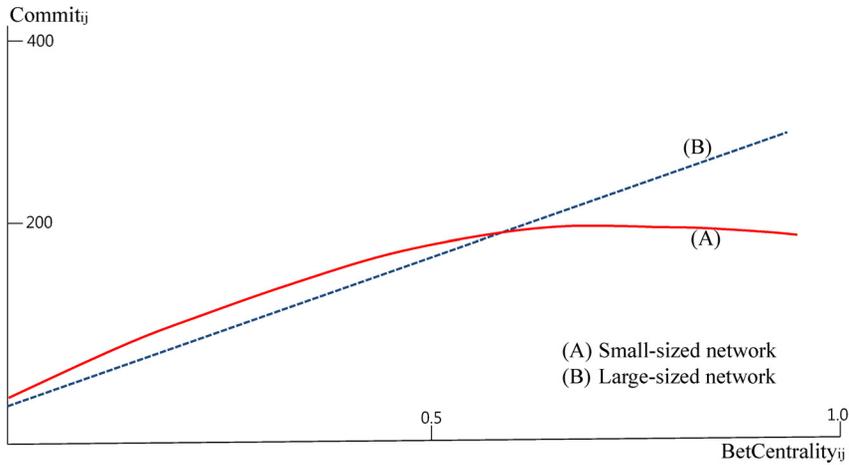
	Small-sized network	Large-sized network	t-value	Moderating effect
	Coefficient (SE) N = 4,630	Coefficient (SE) N = 3,471		
BetCentrality _{ij} → Commit _{ij}	258.95 (5.85)	329.95 (6.02)	-8.32	Support
Adjusted R ²	0.34	0.47		

※ Using the Chin's t-test.

<Table 8> Analysis Results of Moderating Effects : Network Size (Second Order Equation)

	Small-sized network	Large-sized network	t-value	Moderating effect
	Coefficient (SE) N = 4,630	Coefficient (SE) N = 3,471		
BetCentrality _{ij} → Commit _{ij}	523.90 (19.84)	408.20 (23.66)	3.76	Support
BetCentrality _{ij} ² → Commit _{ij}	-335.97 (24.08)	-107.27 (31.36)	-5.88	Support
Adjusted R ²	0.37	0.47		

※ Using the Chin's t-test.



〈Figure 2〉 Moderating Effects of Network Size

석 모델 4)을 대상으로 동일한 분석을 실시하였다. 특이할만한 점은, 매개중심성과 개발자 커밋수와의 관계가 네트워크 사이즈가 작은 경우에는 2차항 모형이, 네트워크 사이즈가 큰 경우에는 1차항 모형이 적합함을 확인할 수 있었다(〈Figure 2〉 참조).

비록 본 연구를 위해 수집된 데이터는 시계열 데이터가 아니기 때문에 오픈소스 프로젝트의 진화에 대한 분석을 하기는 어려우나, 수집된 리파지토리의 네트워크 사이즈의 상이함으로 인해 본 데이터를 통해 네트워크의 성장에 따른 특성을 살펴볼 수 있다. 본 분석(1차항 분석)을 통해 네트워크가 진화할수록(네트워크의 크기가 커질수록[3]) 매개중심성이 큰 개발자의 역할이 더욱 커짐을 확인할 수 있다.

4.2.2 협업 네트워크 내 허브 존재여부

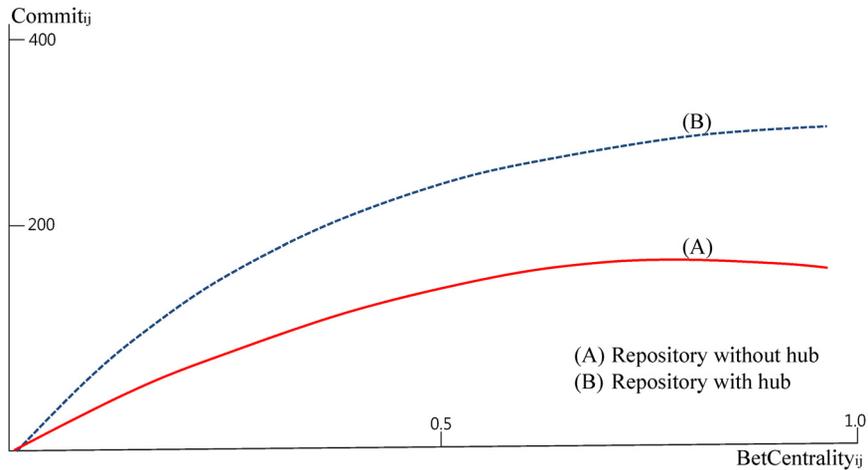
협업 네트워크 내에서 연결중심성이 평균값보다 3표준편차 이상인 노드(개발자)를 허브로 가정하였다. 즉 협업 네트워크 내에서의 허브 개발자는 일반적인 개발자들과는 달리, 다

른 개발자와의 협업이 빈번히 이루어진다. 분석한 결과, 782개의 리파지토리에서의 평균 허브 개수는 0.96개로 1개가 미쳐 되지 않았다. 본 분석에서는 허브를 가지는 리파지토리와 그렇지 않은 리파지토리를 나누어 조절효과를 살펴보았다. 분석 결과, 허브수가 많은 리파지토리의 경우 매개중심성과 커밋수와의 관련성이 더욱 강해짐을 확인할 수 있다. 〈Table 9〉는 1차항 모형(위계적 회귀분석 모델 3)을 대상으로 두 가지 그룹(허브 존재 vs. 허브 미존재)에서의 매개중심성과 개발자 커밋수 관계에 있어서의 차이점을 분석하였으며, 〈Table 10〉에서는 2차항 모형(위계적 회귀분석 모델 4)을 대상으로 동일한 분석을 실시하였다. 매개중심성과 개발자 커밋수와의 관계는 허브 존재여부와 상관없이 2차항 모형이 모두 적합하며, 허브가 존재하는 리파지토리에서의 매개중심성과 개발자 커밋수와의 관련성이 확연하게 커짐을 확인할 수 있다(〈Figure 3〉 참조).

앞서 말한 바와 동일한 맥락에서, 네트워크 크기가 진화할수록 리파지토리 내의 허브가 생

겨날 확률은 높아지게 된다[3]. 이러한 측면에서 살펴봤을 때, 네트워크가 진화할수록(허

브가 존재할수록) 매개중심성이 큰 개발자의 역할이 더욱 커짐을 또 한 번 확인할 수 있다.



〈Figure 3〉 Moderating Effects of Hub

〈Table 9〉 Analysis Results of Moderating Effects : Hub (First Order Equation)

	Repository without hub	Repository with hub	t-value	Moderating effect
	Coefficient (SE) N = 6,904	Coefficient (SE) N = 1,197		
BetCentrality _{ij} → Commit _{ij}	268.46 (4.73)	439.84 (9.48)	-14.25	Support
Adjusted R ²	0.34	0.64		

※ Using the Chin's t-test.

〈Table 10〉 Analysis Results of Moderating Effects : Hub (Second Order Equation)

	Repository without hub	Repository with hub	t-value	Moderating effect
	Coefficient (SE) N = 6,904	Coefficient (SE) N = 1,197		
BetCentrality _{ij} → Commit _{ij}	503.10 (16.41)	694.05 (38.17)	-4.49	Support
BetCentrality _{ij} ² → Commit _{ij}	-304.03 (20.39)	-339.84 (49.49)	0.67	Non-Support
Adjusted R ²	0.36	0.66		

※ Using the Chin's t-test.

4.3 운영자와 비운영자의 네트워크 중심 구조 비교분석

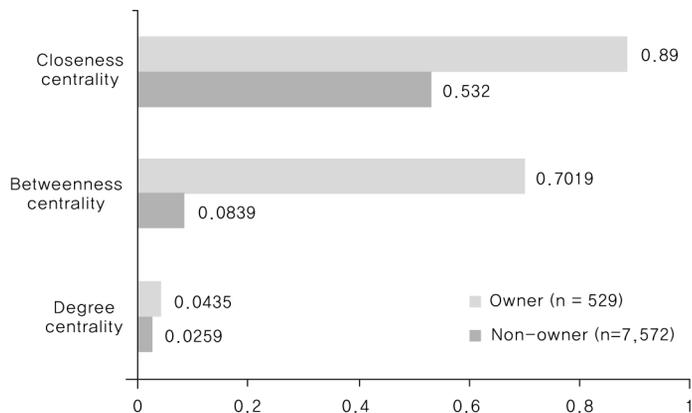
앞서 기여율이 높은 개발자들의 네트워크 중심구조에 대해 초점을 맞추었다면, 본 분석에서는 리파지토리 운영자(owner)의 네트워크 중심구조를 비운영자와의 네트워크 중심구조와 비교·분석하였다. 아래 <Table 11> 및 <Figure 4>에서 제시하는 바와 같이, 운영자의

연결·매개·근접 중심성 모두 비운영자에 비해 통계적으로 유의한 수준에서 높은 수치를 가지는 것으로 나타났다. 특이할 만한 점은, 연결중심성과 근접중심성의 경우 운영자의 중심성이 비운영자에 비해 2배 이상 크지 않은 반면, 매개중심성은 운영자의 경우 비운영자에 비해 8배 이상 높은 것으로 나타났다. 기여율이 높은 개발자와 비슷하게 운영자의 경우에도 매개중심성이 상당히 높은 것으로 나타났다.

<Table 11> Comparison of Network Centrality of Subsamples : Owners versus Non-owners

Variable	Non-owner (n = 7,572) Mean (SD)	Owner (n = 529) Mean (SD)	Total (n = 8,101) Mean (SD)	t-value	p-value
DegCentrality _{ij}	0.0259 (0.0360)	0.0435 (0.0412)	0.0360 (0.0367)	-10.75	0.000
BetCentrality _{ij}	0.0839 (0.1863)	0.7019 (0.2471)	0.1243 (0.2444)	-71.99	0.000
CloCentrality _{ij}	0.5320 (0.1849)	0.8900 (0.1364)	0.5554 (0.2025)	-43.71	0.000

※ Using the t-test.



<Figure 4> Comparison of Network Centrality of Subsamples : Owners versus Non-owners

5. 토론 및 결론

개방형협업의 성공을 연구하기 위해 본 연구에서는 오픈소스 프로젝트의 대표적 플랫폼인 깃허브의 상호협력 네트워크 구조를 사회연결망이론을 통해 살펴보았다. 특히 본 연구에서는 공동연구 네트워크 내에서의 개발자의 역할과 개발자의 자발적 프로젝트에서의 기여도와의 관련성을 살펴보고자 하였다. 이를 위해 대표적인 오픈소스 프로젝트 플랫폼인 깃허브에서 활발한 활동이 이루어지고 있는 782개의 리파지토리를 대상으로 개발자간 협업 네트워크를 분석하였다. 분석을 통해 크게 다음의 세 가지 분석 결과를 얻을 수 있었다. 첫째, 연결·매개·근접 중심성 세 가지 모두 개발자의 커밋수와 긍정적인 관계를 가지는 것으로 나타났다. 또한 이러한 세 가지 관련성 모두 1차항이 아닌 2차항으로 설명할 경우 설명력이 더욱 증가하는 것으로 나타났다. 특히, 다수의 공동개발자를 연결해주는 역할을 수행하는 개발자의 매개중심성이 높을수록 그 개발자의 커밋수 또한 증가하는 경향이 강한 것으로 나타났다. 일반적으로 다수의 개발자와 협업을 많이 한 개발자(연결중심성이 높은 개발자)가 리파지토리 내에서의 커밋수가 많을 것으로 기대하는 경향이 있으나, 본 분석 결과 매개중심성에 의한 설명력이 연결중심성에 의한 설명력보다 확연하게 높은 것으로 나타났다. 이러한 결과는 Abbasi et al.[2]이 철강 산업 관련 연구논문의 공동연구자 네트워크를 통해 규명한 바와 비슷한 결과라 판단할 수 있다. 그들의 연구에 따르면, 특정 연구 도메인의 연구자 네트워크 내에서 매개중심성이 높은 연구자들(즉,

정보의 흐름을 제어할 수 있는 역할을 수행하는 연구자)의 경우 새로운 연구자들과의 협력이 잘 이루어짐을 확인하였다. 비록 그들의 연구에서 종속변수로 본 새로운 연구자와의 협업 여부는 본 연구의 종속변수인 개발자의 기여도와는 차이를 보이나, 공동연구 연결망에서 매개중심성이 높은 개발자(연구자)들이 공동연구에 있어 주요한 역할을 수행한다는 점에서 동일한 맥락이라 볼 수 있다. Cross and Cummings[7] 또한 지식중심업무에 있어 정보 네트워크 내에서의 매개중심성이 높은 사원이 문제를 효과적으로 해결하기 위해 필요한 정보를 얻기가 용이해져 성과 또한 높아짐을 밝힌 바 있다.

둘째, 앞서 리파지토리의 속성에 따라 개발자의 매개중심성과 개발자의 커밋수와의 관계가 어떻게 달라지는지 알아보기 위해 리파지토리의 네트워크 사이즈 및 허브 존재유무의 조절효과를 분석해보았다. 분석 결과, 네트워크 사이즈가 클수록, 네트워크 내에 허브가 존재할수록 매개중심성과 개발자 커밋수의 관련성이 커지는 것으로 나타났다. 오픈소스 프로젝트는 자발적 조직의 형태로서, 운영자가 운영을 시작한 이후 시기가 지날수록 네트워크가 확대되는 형태로 나타나게 된다. 리파지토리 운영초기에는 개발자간 상호협력보다는 소수의 사람들, 특히 운영자에 의해 운영이 이루어지다가 점점 리파지토리가 활성화될수록 개발자간 네트워크가 성장하게 된다. Barabási[3]에 따르면, 네트워크는 진화될수록 네트워크 사이즈가 커지며, 멱함수 법칙을 따르게 되어 다른 노드에 비해 많은 수의 링크를 가지는 허브가 생겨나게 된다. 본 분석에 따르면, 네트워크가 진화될수록(네트

워크 사이즈가 커지거나, 허브가 생겨나기 시작하면서) 매개중심성의 역할이 더욱 중요해 진다고 볼 수 있다.

셋째, 추가적으로 리파지토리의 운영자와 비운영자의 네트워크 중심구조를 t검정을 통해 분석한 결과, 운영자의 경우 연결·매개·근접 세 가지 네트워크 중심성 모두 큰 것으로 나타났다. 주목할만한 점은 다른 네트워크 중심성에 비해 매개 중심성에 있어서의 차이가 확연하게 큰 것으로 나타났다. 본 분석 결과는 오픈소스 프로젝트에서 리파지토리 운영자들에게는 다양한 주제로 모듈화된 개발자들을 연결해주는 역할이 중요함을 암시한다.

본 연구의 기여점은 다음과 같이 정리된다. 첫째, 사회연결망이론을 바탕으로 오픈소스 프로젝트의 자발적인 협업 네트워크의 특성을 통해 개방형협업의 성공 가능성을 설명하고자 시도하였다는 점에서 이론적 기여점을 찾을 수 있다. 특히, 기존 연구가 오픈소스 프로젝트의 네트워크(프로젝트) 수준에서의 특성을 통해 프로젝트의 성공여부에 초점을 맞추어 접근을 시도한 반면, 본 연구에서는 프로젝트 내의 개발자를 대상으로 네트워크 중심구조를 도출함으로써 활발한 기여를 하는 개발자의 네트워크 중심구조를 살펴보았다는 점에서 기존 연구와의 차별화를 꾀할 수 있다.

둘째, 본 연구는 기존의 사회과학 연구에서의 방법론에서 다소 벗어나 설문조사가 아닌 실제데이터를 통한 연구라는 점과 개인의 특성뿐만 아니라 개인간의 상호작용에 의한 패턴을 살펴보고자 하는 점에서 차별화를 꾀할 수 있다. 즉, 최근 각광받고 있는 빅데이터의 일련인 웹데이터를 자동으로 수집·분석하여 사회 현상을 살펴보았다는 점과 개발자·리파

지토리 자체가 가지는 특성뿐만 아니라 개발자간의 상호협력 관계 또한 분석의 대상으로 포함한 점 등을 기존연구와의 차이점으로 볼 수 있다. 기존 깃허브를 연구도메인으로 하는 연구[9, 36]의 경우 깃허브 실제데이터를 바탕으로 한 기술적 통계분석 부분에 치우쳐져 있는 반면, 본 연구에서는 깃허브 개발자간 협업 네트워크를 통해 기여도가 높은 개발자를 살펴 보았다는 데에서도 그 의의를 찾을 수 있다.

개방형협업의 중요성이 날로 커지고 있는 현 상황에서 깃허브 오픈소스 프로젝트의 성공을 위한 개발자의 기여도에 대한 분석은 시사하는 바가 클 것으로 기대된다. 본 연구는 오픈소스 프로젝트뿐만 아니라 더 나아가 개방형협업의 형태로 이루어질 다양한 프로젝트의 성공을 위한 효율적인 운영에 있어 가이드 라인을 제공해줄 수 있을 것으로 기대된다. 본 연구를 통해 개방형협업의 성공을 위해서는 매개중심성이 높은 개발자를 중점적으로 관리할 필요성이 커짐을 확인하였다. 특히, 협업 네트워크가 진화할수록 매개 중심성을 가지는 개발자의 중요성은 더욱 커지는 것으로 나타났다. 따라서, 이러한 개방형협업의 효율적 운영을 위해서는, 매개 중심성이 높은 개발자에게 특정 권한을 부여하는 등의 정책이 필요할 것으로 판단된다. 예를 들면, 아마존닷컴은 만족도가 높은 리뷰를 많이 쓴 리뷰어에게 디지털 배지(badge)를 부여함으로써 질 높은 리뷰의 양을 늘리고자 노력하고 있으며, 세계적인 소셜 뉴스 사이트인 허핑턴 포스트 또한 트위터, 슈퍼유저, 모더레이터 등의 배지를 부여함으로써 방문하는 사람들의 적극성을 높이고, 서로간의 소통을 강화하기 위한 전략을 활용하고 있다. 배지 시스템 등의 사례에서 보는

바와 같이 매개중심성이 높은 개발자들에게 특정한 권한을 부여하는 것은 기여도가 높은 개발자들에게 외적 동기(external motivation)로 작용함으로써 협업의 효율성을 증가시켜 줄 것으로 기대된다.

본 연구에서 더 나아가 리파지토리의 진화 과정에 대한 시계열 데이터를 수집함으로써 공동연구 네트워크 진화 형태를 살펴보고, 프로젝트의 성공을 위해 중요한 역할을 수행하는 기여자들의 네트워크 중심구조 변화를 살펴보는 것도 향후 연구로서 의미가 있을 것으로 판단된다.

References

- [1] Abbasi, A., Altmann, J., and Hossain, L., "Identifying the effects of co-authorship networks on the performance of scholars : A correlation and regression analysis of performance measures and social network analysis measures," *Journal of Informetrics*, Vol. 5, No. 4, pp. 594-607, 2011.
- [2] Abbasi, A., Hossain, L., and Leydesdorff, L., "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks," *Journal of Informetrics*, Vol. 6, No. 3, pp. 403-412, 2012.
- [3] Barabási, A. L., *Linked : How everything is connected to everything else and what it means for business, science, and everyday life* : Basic Books, 2014.
- [4] Borgatti, S., "Centrality and AIDS," *Connections*, Vol. 18, No. 1, pp. 112-114, 1995.
- [5] Burt, R. S., *Structural holes : The social structure of competition*, Cambridge, MA : Harvard University Press, 1995.
- [6] Cho, H., Gay, G., Davidson, B., and Ingraffea, A., "Social networks, communication styles, and learning performance in a CSCL community," *Computers and Education*, Vol. 49, No. 2, pp. 309-329, 2007.
- [7] Cross, R. and Cummings, J. N., "Tie and network correlates of individual performance in knowledge-intensive work," *Academy of Management Journal*, Vol. 47, No. 6, pp. 928-937, 2004.
- [8] Crowston, K. and Howison, J., "The social structure of free and open source software development," *First Monday*, Vol. 10, No. 2, 2005.
- [9] Dabbish, L., Stuart, C., Tsay, J., and Herbsleb, J., "Social coding in github : transparency and collaboration in an open software repository," *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM 2012, pp. 1277-1286, 2012.
- [10] Freeman, L. C., "Centrality in social networks conceptual clarification," *Social Networks*, Vol. 1, No. 3, pp. 215-239, 1979.
- [11] Gargiulo, M., Ertug, G., and Galunic, C., "The two faces of control : Network clo-

- sure and individual performance among knowledge workers,” *Administrative Science Quarterly*, Vol. 54, No. 2, pp. 299–333, 2009.
- [12] Gousios, G. and Spinellis, D., “GHTorrent : Github’s data from a firehose,” Paper presented at the Mining Software Repositories (MSR), 2012 9th IEEE Working Conference on, 2012.
- [13] Grewal, R., Lilien, G. L., and Mallapragada, G., “Location, location, location : How network embeddedness affects project success in open source systems,” *Management Science*, Vol. 52, No. 7, pp. 1043–1056, 2006.
- [14] Huang, M., Ahn, J., and Jahng, J., “Patterns of Collaboration Networks : Coauthorship Analysis of MIS Quarterly from 1996 to 2004,” *The Journal of Society for e-Business Studies*, Vol. 13, No. 4, pp. 193–207, 2008.
- [15] Kidane, Y. H. and Gloor, P. A., “Correlating temporal communication patterns of the Eclipse open source community with performance and creativity,” *Computational and mathematical organization theory*, Vol. 13, No. 1, pp. 17–27, 2007.
- [16] Kim, Y. H., *Social Network Analysis*, Parkyoungsa, Seoul, 2007.
- [17] Kim, Y. T., “Analysis of the employment channels of disabled employees : Focus on social network,” *Disability and Employment*, Vol. 21, No. 3, pp. 5–26, 2011.
- [18] Krishnamurthy, S., “Cave or community? : An empirical examination of 100 mature open source projects,” *First Monday*, Vol. 7, No. 6, 2002.
- [19] Lee, J. K., *Open collaboration*, INITIAL COMMUNICATIONS Corp, Seoul, 2012.
- [20] Lee, J. and Kim, S., “Exploring the role of social networks in affective organizational commitment : Network centrality, strength of ties, and structural holes,” *The American Review of Public Administration*, Vol. 41, No. 2, pp. 205–223, 2011.
- [21] Madey, G., Freeh, V., and Tynan, R., “The open source software development phenomenon : An analysis based on social network theory,” *AMCIS 2002 Proceedings*, p. 247, 2002.
- [22] Mitchell, J. C., *The concept and use of social networks*, Bobbs-Merrill, 1969.
- [23] Mizruchi, M. S., Stearns, L. B., and Fleischer, A., “Getting a bonus : social networks, performance, and reward among commercial bankers,” *Organization Science*, Vol. 22, No. 1, pp. 42–59, 2011.
- [24] Newman, M. E., “The structure and function of networks,” *Computer Physics Communications*, Vol. 147, No. 1, pp. 40–45, 2002.
- [25] Oh, W. and Jeon, S., “Membership herding and network stability in the open source community : The Ising perspective,” *Management Science*, Vol. 53, No. 7, pp. 1086–1101, 2007.
- [26] Ortiz, M. R., Hoyos, J. R. C., and López, M. G. R., “The social networks of aca-

- demic performance in a student context of poverty in Mexico,” *Social networks*, Vol. 26, No. 2, pp. 175–188, 2004.
- [27] Park, J. H., Cho, Y. H., and Kim, J. K., “Social network : A novel approach to new customer recommendations,” *Journal of Intelligence and Information Systems*, Vol. 15, No. 1, pp. 123–140, 2009.
- [28] Pil, F. K. and Leana, C., “Applying organizational research to public school reform : The effects of teacher human and social capital on student performance,” *Academy of Management Journal*, Vol. 52, No. 6, pp. 1101–1124, 2009.
- [29] Riehle, D., Ellenberger, J., Menahem, T., Mikhailovski, B., Natchetoi, Y., Naveh, B., and Odenwald, T., “Open collaboration within corporations using software forges,” *Software, IEEE*, Vol. 26, No. 2, pp. 52–58, 2009.
- [30] Scott, J., “Social network analysis,” *Sociology*, Vol. 22, No. 1, pp. 109–127, 1988.
- [31] Scott, J., *Social network analysis : A handbook*, Sage, 1991.
- [32] Simon, H. A., *The architecture of complexity*, Springer, 1991.
- [33] Sin, H. P., Nahrgang, J. D., and Morgeson, F. P., “Understanding why they don’t see eye to eye : An examination of leader-member exchange (LMX) agreement,” *Journal of Applied Psychology*, Vol. 94 No. 4, pp. 1048–1057, 2009.
- [34] Singh, P. V., Tan, Y., and Mookerjee, V., “Network Effects : The Influence of Structural Capital on Open Source Project Success,” *MIS Quarterly*, Vol. 35, No. 4, pp. 813–829, 2011.
- [35] Sparrowe, R. T., Liden, R. C., Wayne, S. J., and Kraimer, M. L., “Social networks and the performance of individuals and groups,” *Academy of management Journal*, Vol. 44, No. 2, pp. 316–325, 2001.
- [36] Takhteyev, Y. and Hilt, A., “Investigating the geography of open source software through GitHub,” *Working Paper*, 2010.
- [37] Thung, F., Bissyandé, T. F., Lo, D., and Jiang, L., “Network structure of social coding in GitHub,” *Paper presented at the Software Maintenance and Reengineering (CSMR), European Conference on*, 2013.
- [38] Yang, H. L. and Tang, J. H., “Effects of social network on students’ performance : a web-based forum study in Taiwan,” *Journal of Asynchronous Learning Networks*, Vol. 7, No. 3, pp. 93–107, 2003.
- [39] Zhang, X. and Venkatesh, V., “Explaining employee job performance : the role of online and offline workplace communication networks,” *Mis Quarterly*, Vol. 37, No. 3, pp. 695–722, 2013.
- [40] Wikipedia, http://en.wikipedia.org/wiki/GitHub#cite_note-ucc-4, 2014.10.29 access.

저 자 소 개



백현미 (E-mail : lotus1225@hanyang.ac.kr)
2002년 포항공과대학교 화학공학과 (학사)
2004년 한국정보통신대학교(현. 한국과학기술원)
IT 경영학과 (석사)
2013년 서울대학교 경영학과 (박사)
2004년~2013년 한국전자통신연구원 선임연구원
2013년~현재 한양대학교 언론정보대학 정보사회학과 조교수
관심분야 소셜미디어, 온라인구전, 개방형협업, ICT R&D정책 등



오세환 (E-mail : sehwan@kita.net)
2001년 서울대학교 경제학부 (학사)
2007년 카네기멜론대학 MSIT (석사)
2015년 서울대학교 경영학과 (박사)
2003년~현재 한국무역협회 국제무역연구원 수석연구원
관심분야 소셜미디어, 온라인구전, 전자상거래, 전자무역