Densification Power Law 기반 비트코인 네트워크 통계 데이터 분석

백의준, 신무곤, 지세현, Huru Hasanova, 김명섭
고려대학교
{ pb1069, tm0309, sxzer, hhuru, tmskim }@korea.ac.kr

The Analysis of Bitcoin Network Statistical Data Based on Densification Power Law

Ui-Jun Baek, Mu-Gon Shin, Se-Hyun Jee, Huru Hasanova, Myung-Sup Kim
Korea Univ.

요 약

사토시 나카모토에 의해 블록체인 기술이 개발되고 비트코인이 새로운 암호화폐 시장을 개척한 이후 여러 암호화폐들이 등장하고 그 수와 규모는 나날이 증가하고 있다. 또한 블록체인 기술의 익명성과 여러 취약점을 이용한 범죄들이 발생하고 있으며 이에 취약점 개선과 범죄 예방을 위한 많은 연구들이 진행되고 있으나 범죄를 저지르는 사용자들을 탐지해내기엔 역부족이다. 따라서 네트워크 내 자금 세탁, 자금 탈취 등 이상 행위를 탐지 하는 것은 매우 중요하며 이에 본 논문에서는 비트코인 네트워크의 트랜잭션 및 User 그래프의 Feature들을 분석하고 네트워크 내 이상 탐지에 적절한 Feature들을 제시한다.

Ⅰ. 서 론

사토시 나카모토에 의해 블록체인 기술이 개발되고 비트코인이 새로운 암호화폐 시장을 개척한 이후 여러 암호화폐들이 등장하였으며 그 수와 규모는 나날이 증가하고 있다. 블록체인 시장의 급격한 성장에 따라 블록체인 기술의 익명성과 취약점을 이용하는 여러 범죄들이 발생하고 있으며 현재까지도 지속적으로 발생하고 있다. 취약점 개선과 범죄 예방을 위한 많은 연구들이 진행되고 있으나 악성행위를 예방하고 그 행위를 저지르는 악성 사용자들을 정확히 탐지해내기엔 역부족이다. 그러므로 네트워크 내 자금 세탁 및 탈취와 같은 이상 행위를 탐지하는 것은 매우 중요하며 본 논문에서는 비트코인 네트워크의 트랜잭션 및 User 그래프의 Feature들을 분석하고 이상 행위 탐지에 적절한 Feature들을 제시한다.

본 논문은 1장 서론, 2장 관련 연구, 3장 본론, 4장 분석 결과 순으로 설명하고 마지막 5장에서 결론과 향후연구를 제시한다.

Ⅱ. 관련 연구

[1]은 시간에 따른 그래프 변화의 특성을 설명한다. 정상적인 네트워크의 그래프의 노드와 에지 수가 로그 스케일 상에서 선형함수의 형태를 가진다는 Densification Power Law를 제시하는데 이 법칙에 따라 특정 네트워크 의 그래프의 분포가 비선형적일 경우 네트워크 내 이상이 있을 수 있다고 판단할 수 있다.

[2,3]은 블록체인 네트워크로부터 User 데이터를 추출하고 이를 특정기준에 따라 분류하거나 클러스터링 알고리즘을 통해 관련된 여러 지갑들의 연관성을 추출하는 방법을 제안하였다. [2,3] 모두 블록체인 네트워크를 분석하고자 하는 사용자에게 Forensic 분석의 가능성을 제시할 순 있으나 Heuristic한 기준과 수동적인 분석으로 시시각각 변화하는 네트워크의 특성을 모두 반영하기 힘들다는 한계점을 지닌다.

[4]는 비트코인 네트워크로부터 트랜잭션 데이터로부터 User 그래프와 트랜잭션 그래프를 추출하고 군집화하고 각 클러스터 내 이상치를 계산하는 수식을 통해 의심스러운 트랜잭션 혹은 User를 탐지하는 방법을 제안하였다. 그러나 탐지에 사용한 Feature들의 종류가 적어 정확한 탐지가 어렵다는 한계점을 지니며 이에 본 논문에서는 다양한 Feature과 그 통계정보를 Densification Power Law에 따라 분석하고 그 분석 결과를 통해 정확한 이상탐지에 있어 적절한 Feature들을 제시한다.

Ⅲ. 본론

본 장에서는 핵심 개념 및 데이터 수집 및 처리에 대해 설명한다.

i. Power Degree & Densification Power Laws

Densification Power Law는 노드(N)와 에지(E)로 이루어진 그래프에서 특정 시간 t의 노드 개수의 α제곱은 특정시간 t의 에지의 개수에 비례한다는 법칙이며 이는 수식 1과 같다.

$$E(t) \propto N(t)^{a} \quad (1)$$

이를 변형하여, 실제 정상적인 네트워크에서 $P(k)$를 차수 k를 가지는 노드의 Feature라고 정의하고 γ가 양의 정수일 때 $P(k)$는 차수 k의 역수에 비례하며 이는 수식 2와 같다. $P(k)$는 잔액, 트랜잭션 사이즈, 총 거래금액 등으로 대체될 수 있다.

$$P(k) \propto k^{-\gamma} \quad (2)$$

ii. 데이터 수집 및 통계 추출

비트코인 네트워크의 1부터 20만번 째 블록에 담긴 트랜잭션 데이터를 수집하였으며 트랜잭션 데이터로부터 User 데이터를 추출하였다. 이를 노드가 User 데이터인 노드 그래프, 노드가 트랜잭션인 트랜잭션 그래프의 형태로 변형하며 두 그래프는 입력과 출력이 존재하므로 방향성을 가지는 차수(In-Degree, Out-Degree)를 가진다. 마지막으로 그래프로부터 합, 최댓값, 최솟값, 평균, 표준편차를 추출하였으며 추출한 모든 Feature는 그림 [1,2]에 나타나 있다.

그림 1과 같이 User그래프에서 In/Out-Degree의 Number of Degrees 총 2개와 In/Out-Degree 2가지 방향성의 Value/Size/Weight 3개의 Features, 5개의 통계정보 총 30(2*3*5)개를 추출하였다.

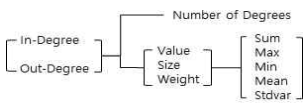

그림 1. User그래프에서 추출한 Feature Set

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(No.2018-0-00539-001,블록체인의 트랜잭션 모니터링 및 분석 기술개발)을 받고 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원(NRF-2018R1D1A1B07045742)을 받아 수행된 연구임

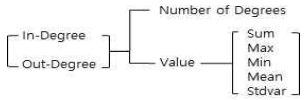

그림 2. 트랜잭션 그래프에서 추출한 Feature Set

그림 2와 같이 트랜잭션 그래프에서 In/Out-Degree의 Number of Degrees 총 2개와 In/Out-Degree 2가지 방향성의 Value, 이에 대한 5개의 통계정보 총 10(2*1*5)개를 추출하였다.

iii. 데이터 그래프 화

수집하고 추출한 데이터를 비교가 용이하도록 로그 스케일 그래프로 나타내었으며 x축에는 공통적으로 차수(In-Degree, Out-Degree)로 설정하고 y축은 추출한 각각의 Feature로 설정하였다.

Ⅳ. 실험 결과

본 장에서는 추출한 데이터를 그래프로 나타내고 이에 대해 설명한다.

서론에서 언급했듯이 그래프의 분포가 비선형일 경우 해당 네트워크 내이상이 있을 수 있다. 따라서 그래프의 분포를 보며 분석하고 비선형적인그래프의 분포를 가지는 Feature를 찾는다.

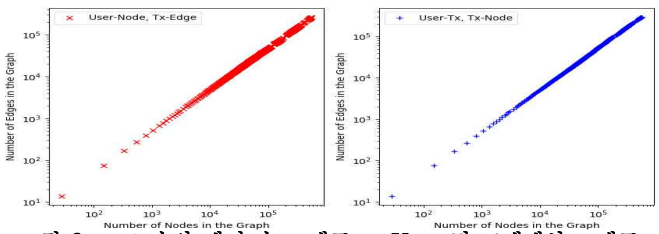

그림 3. 노드 수와 에지 수 그래프 - User 및 트랜잭션 그래프

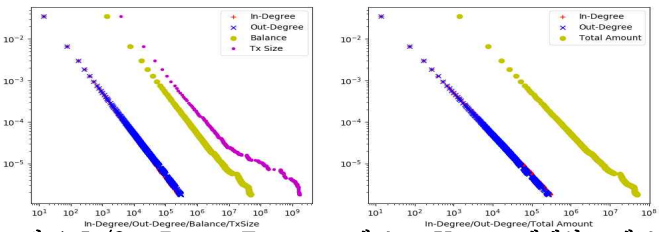

그림 4. In/Out-Degree Feature 그래프 - User, 트랜잭션 그래프

그림 3의 User 및 트랜잭션의 노드-에지 그래프와 [4]에서도 제시했던 그림 4의 In/Out-Degree에서는 단 한 개 Tx Size를 제외하고 모든 그래프가 선형함수의 형태를 나타내는 것을 확인했으며 이러한 Feature들로는 정상과 이상을 구분할 명백한 특징이 없다고 말할 수 있으며 정확한 이상 탐지가 어렵다고 판단된다.

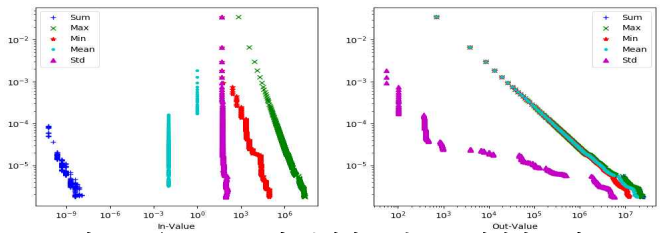

그림 5. In/Out Value의 통계정보 분포-트랜잭션 그래프

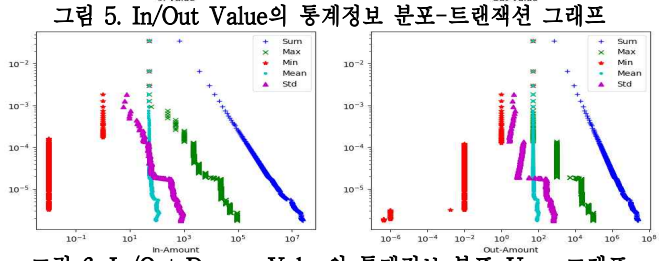

그림 6. In/Out Degree Value의 통계정보 분포-User 그래프

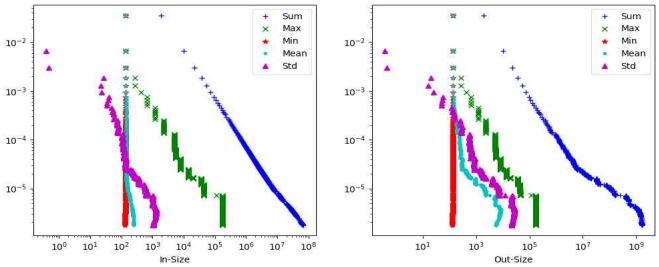

그림 7. In/Out Degree Size의 통계정보 분포-User 그래프

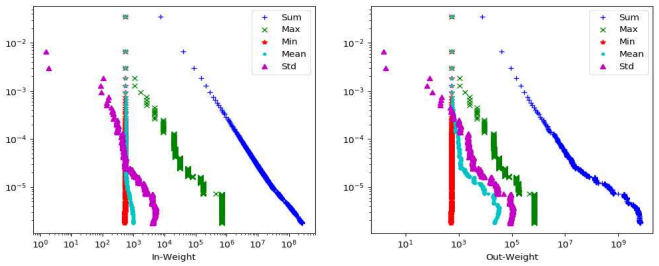

그림 8. In/Out Degree Weight의 통계정보 분포-User 그래프

그림 [5-8]은 User 및 트랜잭션 그래프의 Feature에서 통계정보를 추출하여 Degree-Stat 형태의 그래프를 나타낸 것이며 대부분의 통계정보를 이용한 그래프 분포에서 비선형적인 형태가 나타나는 것을 확인하였으며 5개의 통계정보 중 특히 표준 편차 정보에서 뚜렷한 비선형적인 형태가 나타나는 것을 볼 때 그래프 분석을 통한 이상 탐지에서 표준편차 값이 중요함을 확인하였다.

Ⅴ. 결론

본 논문은 비트코인 네트워크의 트랜잭션 데이터와 User 데이터를 수집 및 추출하고 이들의 통계정보를 노드와 에지로 이루어진 그래프 형태로 변환하고 이를 분석하였다. 분석 결과를 통해 수집할 수 있는 일반적인 정보보다 통계정보가 명확히 구분할 수 있는 분포를 띈다는 것을 확인하였다. 이를 통해 클러스터링과 같은 이상탐지를 위한 심화 분석의 가능성을 제시하였다. 향후 연구로는 K-means 알고리즘을 통해 클러스터링을 진행하고 비트코인 네트워크 내 이상 탐지에 대한 연구할 계획이며 이전 연구와의 비교를 통해 본 연구의 객관성을 갖출 예정이다.

참 고 문 헌

[1] Kalodner, Harry, et al. "BlockSci: Design and applications of a blockchain analysis platform." arXiv preprint arXiv:1709.02489 (2017).

[2] Spagnuolo, Michele, Federico Maggi, and Stefano Zanero. "Bitiodine: Extracting intelligence from the bitcoin network." International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2014.

[3] Leskovec, Jure, Jon Kleinberg, and Christos Faloutsos. "Graph evolution: Densification and shrinking diameters." ACM Transactions on Knowledge Discovery from Data (TKDD) 1.1 (2007): 2.

[4] Pham, Thai, and Steven Lee. "Anomaly Detection in the Bitcoin System-A Network Perspective." arXiv preprint arXiv:1611.03942 (2016)