NPM Ekosisteminde Top N Paket İçin Yönlü Karmaşık Ağ Analizi: Merkeziyet, Risk Skoru ve Robustluk

Yusuf Talha ARABACI

21 Ekim 2025

Özet

Bu çalışma, NPM ekosistemindeki popüler Top N paketi, bağımlılık ilişkilerine göre yönlü bir karmaşık ağ (Dependent → Dependency) olarak modellemekte ve yapısal riskleri merkeziyet metrikleriyle incelemektedir. Veri, her çalıştırmada API'lerden (öncelikle ecosyste.ms; yedek olarak npm registry ve npms.io) çekilmektedir. Ağ, NetworkX ile kurulmakta; in-degree, out-degree ve betweenness merkeziyet metrikleri hesaplanmaktadır. Büyük grafiklerde betweenness hesaplaması örnekleme (k) ile hızlandırılmaktadır. Çalışma ayrıca bileşik bir risk skoru (normalize edilmiş in/out/between ağırlıklı toplamı) ve risk tabanlı robustluk analizi (kritik düğümlerin kaldırılması) önermektedir. Üretilen tüm çıktılar (CSV/MD/JSON ve PNG/SVG görseller) results/ dizininde saklanır.

1 Giriş

Yazılım tedarik zinciri saldırılarında (SSCA), tek bir bağımlılığın ele geçirilmesi geniş çapta zincirleme etkilere yol açabilir. NPM ekosistemi, yoğun bağımlılık ilişkilerine sahip olup, paketlerin yapısal konumuna göre sistemik risk taşıyabilmektedir. Bu çalışma, Top N paket üzerinden inşa edilen yönlü bağımlılık ağı ile aşağıdaki sorulara odaklanır:

- Hangi düğümler (paketler) yapısal olarak kritik (yüksek in-degree, yüksek betweenness)?
- Hangi düğümler geniş bağımlılık yüzeyine sahip (yüksek out-degree)?
- Merkeziyetlere dayalı bileşik bir risk skoru ile risk liderleri nasıl sıralanır?
- Kritik düğümler kaldırıldığında ağın bağlanırlığı nasıl değişir (robustluk)?

2 Çalışmanın Amacı

Bu çalışmanın temel amacı, yazılım tedarik zinciri güvenliğini yapısal bir bakış açısıyla yeniden tanımlamak ve mevcut güvenlik değerlendirme yaklaşımlarına ağ bilimi temelli bir ölcüt kazandırmaktır. Geleneksel sistemler (ör. CVSS), riski yalnızca paket içi zafiyetlerle ölçerken; gerçekte risk, paketin bağımlıolduğu ve kendisine bağımlıolan paketlerle kurduğu ilişkilerden de kaynaklanır. Bu nedenle NPM ekosistemindeki paketleri bir *karmaşık ağ* olarak modelleyerek, her bir paketin ağ içindeki yapısal önemini, ele geçirilmesi durumunda yaratabileceği basamaklanma (*cascading*) etkisini ve bunun sistemik güvenlik riski üzerindeki nicel etkilerini bilimsel metriklerle ölçmeyi ve öngörmeyi hedefliyoruz.

3 Veri ve Yöntem

Top N paket listesi, her çalıştırmada API'lerden çekilir. Öncelik ecosyste.ms üzerinde indirmeye göre sıralı paket adlarındadır; başarısız durumlarda npm registry araması (popularity) ve npms.io

(popülerlik skoru) yedek olarak kullanılır. Bağımlılıklar, npm registry'de paketlerin en güncel sürümlerinin `dependencies` alanından okunur; isteğe bağlı olarak `peerDependencies` de dahil edilebilir. Yönlü ağ, Dependent → Dependency yönüyle kurulur. Büyük graflarda betweenness merkeziyeti örneklemeli (k) hesaplanır.

3.1 Parametreler

Varsayilanlar: Top N=200 (defterde 1000 de denenmistir), `include_peer_deps`=False, betweenness icin `sample_k` *otomatik* ($n \leq 1200$ ise tam, aksi halde $k = 200$). Risk agirliklari: $w_{in} = 0.5$, $w_{out} = 0.2$, $w_b = 0.3$. Ciktilar `results/` dizinine yazilir.

4 Ağ Modeli ve Metrikler

Model, NetworkX ile kurulmuş `DiGraph` yapısıdır. Temel metrikler:

- **In-Degree:** Düğüme gelen kenar sayısı (bu pakete dayanan paket sayısı). Ele geçirilirse etki alanını gösterir.
- **Out-Degree:** Düğümün dış bağımlılık sayısı. Bağımlılık zinciri uzunluğu/karmaşıklığına işaret eder.
- **Betweenness:** En kısa yollardaki aracılık. Köprü rolünü ve tek hata noktası potansiyelini gösterir.

5 Bulgular

Bu bölümde, results/ dizinindeki çıktıları kullanarak görsel ve tablolu özetler sunulmaktadır.

5.1 Ağ Görselleştirmeleri

network_full_topN.png bulunamadı

Şekil 1: Top N + bağımlılıkların oluşturduğu yönlü ağ (düğüm boyutu: in-degree, renk: Top N turuncu / diğerleri mavi).

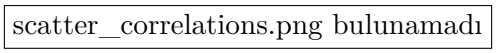
network_topN_only.png bulunamadı

Şekil 2: Sadece Top N düğümlerin indüklenmiş alt-ağı.

5.2 Derece Dağılımları ve Korelasyonlar

degree_histograms.png bulunamadı

Şekil 3: In-Degree ve Out-Degree histogramları (log ölçek).



Şekil 4: Korelasyonlar: In-Degree vs Betweenness (solda), In-Degree vs Out-Degree (sağda).

Şekil 5: İlk 10 lider (In/Out-Degree ve Betweenness).

5.3 Liderler (İlk 10)

5.4 Ek Görseller

5.5 Risk Skoru ve Robustluk

Edge Betweenness İlk 10. Aşağıdaki tablo, en yüksek edge betweenness değerine sahip 10 kenarı göstermektedir.

5.6 Ek Tablolar

Top 20 In-Degree (Tüm Düğümler). metrics_top20_in_degree.tex bulunamadı

Top 20 Risk Skoru. risk_scores_top20.tex bulunamadı

5.7 Basamaklanma (Cascading Impact)

Risk liderleri icin hesaplanan basamaklanma etkisi asagidaki gorselde sunulmustur.

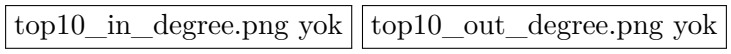
Risk–Basamaklanma Çizimi. Ek olarak, risk skoru ile basamaklanma etkisi arasındaki ilişki de gözlemlenmiştir.

6 Tum Ciktilar ve Dosya Ozeti

6.1 Gorseller (PNG/SVG)

Bu calismada uretilen temel gorseller asagidadir; tamamıresults/ dizinindedir.

- network_full_topN.(png|svg): Top N + bagimliliklarin tam agi
- network_topN_only.(png|svg): Sadece Top N indüklenmis alt-ag
- degree_histograms.(png|svg): In/Out-degree histogramlari (log olcek)
- scatter_correlations.(png|svg): In-Degree vs Betweenness; In-Degree vs Out-Degree
- top10_in_degree.(png|svg), top10_out_degree.(png|svg), top10_betweenness.(png|svg), top10_leaders.(png|svg)
- top20_risk.(png|svg): Bilesik risk skoruna gore ilk 20
- cascade_impact_top20.(png|svg): Basamaklanma etkisi (risk liderleri)
- risk_vs_cascade.(png|svg): Risk ile basamaklanma iliskisi (scatter)



Şekil 6: İlk 10 In-Degree (sol) ve Out-Degree (sağ).

Şekil 7: İlk 10 Betweenness.

6.2 Veri ve Tablolar

Asagidaki dosyalar metrik ve arakatman ciktilarini icerir:

Dosya	Aciklama
`edges.csv`	Kenar listesi (source=dependent, target=dependency)
`metrics.csv`	Dugum metrikleri (in/out/between, is_topN)
`risk_scores.csv`	Bilesik risk skorları
`edge_betweenness_top10.csv`	En yuksek edge betweenness 10 kenar
`robustness_risk.json`	Risk tabanli kaldirma senaryolari icin baglanirlik
`graph_stats.json`	Genel ag istatistikleri (dugum/kenar, bilesen, cap)
`top_packages.txt`	Kullanilan Top N paket adlari
`report.md`	Kisa siralama raporu (ilk 20 listeler)
`metrics_top20_in_degree.tex`	Top 20 In-Degree (LaTeX tablo)
`risk_scores_top20.tex`	Top 20 Risk Skoru (LaTeX tablo)
`cache_deps.json`	Bagimlilik sorgulari icin onbellek

7 Risk Skoru Tanımı

Normalize edilmiş metrikler (min-max) üzerinden ağırlıklandırılmış bir risk skoru tanımlanmıştır:

$$\text{risk}(n) = w_{in}\,\tilde{d}_{in}(n) + w_{out}\,\tilde{d}_{out}(n) + w_b\,\tilde{b}(n),$$

burada $\tilde{d}_{in}, \tilde{d}_{out}, \tilde{b}$ sırasıyla in-degree, out-degree ve betweenness'in normalize edilmiş değerleridir. Ağırlıklar (ör. $w_{in} = 0.5, w_{out} = 0.2, w_b = 0.3$) senaryoya göre ayarlanabilir.

8 Robustluk Analizi

Risk skoruna göre en kritik $k \in \{1, 3, 5\}$ düğüm kaldırıldığında zayıf bağlanırlık bileşen sayısı, en büyük bileşen boyutu ve (mümkünse) çap raporlanır. Ayrıntılar results/ dizinindeki `robustness_risk.json` dosyasındadır.

9 Sınırlamalar ve Gelecek Çalışmalar

Sınırlamalar: (i) Varsayılan olarak yalnız `dependencies` kullanılır; `peerDependencies` isteğe bağlıdır. (ii) Betweenness hesaplaması büyük graflarda örneklemelidir; kesin değerlerin yakınsaması ağı büyüklüğüne ve k seçimine bağlıdır. (iii) Global dependent sayıları doğrudan dahil edilmemiştir.

Gelecek Çalışmalar: (i) Global dependent metriklerinin entegrasyonu, (ii) ağırlıkların veri odaklı (öğrenilmiş) belirlenmesi, (iii) çok katmanlı ağ modelleme (örn. dev/peer/optional bağımlılık katmanları), (iv) GEXF/GraphML çıktılarıyla etkileşimli görselleştirme.

top20_risk.png bulunamadı

Şekil 8: Bileşik risk skoruna göre ilk 20 paket.

cascade_impact_top20.png bulunamadı

Şekil 9: Risk skoruna göre ilk 20 paket için basamaklanma (cascading impact) büyüklüğü.

10 Sonuç

Bu çalışma, NPM ekosistemindeki popüler paketler için yönlü bağımlılık ağını kullanarak yapısal riskleri nicel olarak analiz etmektedir. Merkeziyet metrikleri, bileşik risk skoru ve robustluk analizleri, kritik düğümlerin ve köprülerin pratikte nasıl belirleneceğine dair net bir çerçeve sunmaktadır.

Çoğaltılabilirlik. Tüm kod ve çıktı üretimi rapoyla birlikte depo içindedir. `analysis.ipynb` defteri çalıştırılarak results/ dizini yeniden üretilebilir ve bu makale LaTeX dosyası ile raporlanabilir.

Kaynaklar

[1] Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

[2] Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*.

[3] Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

risk_vs_cascade.png bulunamadı

Şekil 10: Risk skoru ile basamaklanma etkisi arasındaki ilişki (scatter).