Gráfcsomópontok kockázat alapú osztályozása

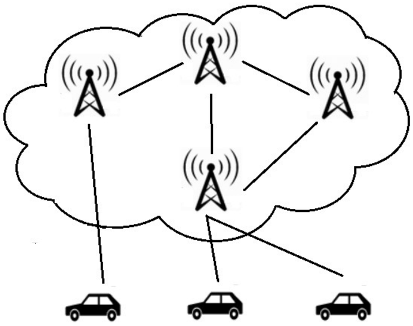
Zentai Dániel*

**Óbudai Egyetem, Mechatronikai és Járműtechnikai Intézet, 1081 Budapest, Népszínház utca 8.*
(e-mail: zentai.daniel@bgk.uni-obuda.hu).

Absztrakt: Napjainkban a vezeték nélküli technológiák elterjedésével együtt dinamikusan fejlődnek **a** felszíni közlekedés kommunikációs technológiái. A felszíni közlekedés kommunikációs hálózatait két fő komponensre bonthatjuk. Az egyik a járművek közötti ad hoc hálózat (Vehicular ad hoc network, VANET), a másik pedig a VANET-et támogató útmenti berendezésekből álló, fix csomópontokkal rendelkező hálózat. Ebben a dolgozatban a hálózat fix elemeinek kockázati osztályokba sorolására javaslunk egy algoritmust.

1. BEVEZETÉS

A felszíni közlekedés kommunikációs hálózatait két fő komponensre bonthatjuk. Az egyik a járművek közötti ad hoc hálózat (Vehicular ad hoc network, VANET), a másik pedig a VANET-et támogató útmenti berendezésekből álló, fix csomópontokkal rendelkező hálózat (S. I. Boucetta, Zs. Cs. Johanyák, L. Pokorádi, 2017).



1. ábra

Ezek közül mi az infrastruktúra fix elemei által alkotott hálózatot fogjuk vizsgálni. Kommunikációs hálózatok matematikai modellezésének természetes eszköze a gráfelmélet. Gráfnak nevezzük csomópontoknak élek által összekötött (általában véges) halmazát. A gráfok megbízhatóságának, támadásokkal szembeni állóképességének egy lehetséges mérőszáma a többszörös összefüggőség. Ez a gráfparaméter mutatja meg, hogy hány élt, vagy csomópontot törölhetünk ki a gráfból oly módon, hogy a megmaradt élek és csomópontok továbbra is egy összefüggő gráfot alkossanak. A Zentai D. (2017) cikkben gráfok többszörös összefüggőségét vizsgáltuk, itt arra térünk ki, hogy ha egy gráf csomópontjai között különböző lokális összefüggőséget követelünk meg, akkor hogyan lehet a csomópontoknak ezen követelményeknek megfelelően osztályokba sorolni.

2. GRÁFELMÉLETI HÁTTÉR

2.1 Alapfogalmak

Ebben a fejezetben gráfelméleti alapfogalmakat ismertetünk, melyek szükségesek a továbbiak megértéséhez. A gráfelmélet alapfogalmainak alaposabb megértéséhez javaslom a Katona Y. Gy, Recski A., Szabó Cs. (2002), illetve a Lovász L., Pelikán J., Vesztergombi K. (2010) könyveket.

Legyen $V = \{v_1, v_2, \ldots, v_n\}$ véges halmaz, és legyen E a V halmaz bizonyos kételemű részhalmazainak egy halmaza, azaz $E \subseteq \binom{V}{2}$. Az ebből a két halmazból álló G = (V,E) rendezett párt véges egyszerű gráfnak nevezzük.

A V = V(G) halmaz elemeit a G gráf csúcsainak, vagy pontjainak, az E = E(G) halmaz elemeit pedig a G gráf éleinek nevezzük. A G = (V,E) gráfban a $v, w \in V$ csúcsokat szomszédosnak nevezzük, ha őket él köti össze, azaz ha $\{v,w\} \in E$. A G = (V,E) gráf irányított gráf abban az esetben, ha minden élének van egy iránya is, azaz megkülönböztetjük a $(v,w) \in E$ élt a $(w,v) \in E$ éltől. Ebben az esetben $E \subseteq \binom{V}{2}$ helyett $E \subseteq V \times V$. A gráfok szokásos geometriai reprezentációjában az éleket irányítatlan esetben szakaszokkal, vagy görbékkel, irányított esetben pedig nyilakkal ábrázoljuk.

A G = (V,E) gráf $v \in V$ csúcsának szomszédjainak a halmazát N(v) a v csúcs fokszámát pedig d(v) jelöli, azaz $d(v)=|N(v)|$. Irányított gráf esetében megkülönböztetjük a v csúcs befokát és kifokát. A v csúcs befoka azon éleknek a száma, melyeknek végpontja v, a v csúcs kifoka pedig azon élek száma, melyeknek a kezdőpontja v.

A séta a gráfban csúcsok és élek váltakozó $v_0e_1v_1\ldots e_kv_k$ sorozata, ahol mindegyik él a sorozatban őt megelőző és őt követő csúcsokat köti össze, azaz $e_i = \{v_{i-1},v_i\}$. A vonal olyan séta, amelyben egy él legfeljebb egyszer szerepelhet, az út pedig olyan vonal, amelyben minden csúcs is maximum egyszer szerepel. A séta, vonal vagy út hosszának az ezek során érintett élek számát nevezzük.

Online: ISBN 978-963-88875-3-5
CD: ISBN 978-963-88875-2-8
Editor: Dr. Péter Tamás

A G = (V,E) gráfot összefüggőnek nevezzük, ha bármely u ∈ V pontjából bármely v ∈ V pontjába vezet u kezdőpontú és v végpontú út. Legyen G = (V,E) nem feltétlenül összefüggő gráf. G ponthalmazának egy C ⊆ V(G) részhalmazát akkor nevezzük összefüggőségi komponensnek, ha teljesül rá, hogy bármely u ∈ C pontból bármely v ∈ C pontba vezet út, de semelyik C-beli pontból nem vezet út a V \ C halmaz semelyik pontjába sem. A G gráf összefüggőségi komponenseinek számát c(G) jelöli. Egy G gráf tehát pontosan akkor összefüggő, ha c(G) = 1.

2.2 Többszörös összefüggőség

Hálózatok topologikus értelemben vett hibatűrésének egy természetes mérőszáma a hálózatot modellező gráf többszörös összefüggősége.

Definíció: Azt mondjuk, hogy a G = (V,E) gráf k-szorosan él-összefüggő, vagy röviden k-él-összefüggő, ha G-nek legalább k+1 pontja van (azaz |V(G)| ≥ k+1), és bárhogyan hagyunk el G-ből legfeljebb k darab élt, a kapott G' gráf összefüggő marad. A legnagyobb olyan k értéket, ami a fenti feltételeket teljesíti, a gráf él-összefüggőségi számának nevezzük, és λ(G)-vel jelöljük.

Definíció: Azt mondjuk, hogy a G = (V,E) k-szorosan összefüggő, vagy röviden k-összefüggő, ha G-nek legalább k+1 pontja van (azaz |V(G)| ≥ k+1), és bárhogyan hagyunk el G-ből legfeljebb k darab csúcsot, a kapott G' gráf összefüggő marad. A legnagyobb olyan k értéket, ami a fenti feltételeket teljesíti, a gráf összefüggőségi számának nevezzük, és κ(G)-vel jelöljük.

A fentiek közül a k-él-összefüggőség modellezi az összeköttetések, a k-összefüggőség pedig a csomópontok támadásával szemben támasztott megbízhatósági követelményt. Algoritmikusan könnyedén kiszámolható egy gráf él-összefüggőségi, illetve összefüggőségi száma. Erre szolgálnak a folyamalgoritmusok (Jordán T., Recski A., Szeszlér D., 2004). Közvetlenül a definíció alapján nem tudnánk hatékonyan kiszámolni egy gráf összefüggőségi számát, így a folyamalgoritmusok sem ezt a módszert követik, hanem a definíció egy ekvivalens átfogalmazását használják. A következő két tétel Menger tételének közvetlen következményei.

Tétel: Az alábbiak ekvivalensek:

- A G = (V,E) gráf k-szorosan él-összefüggő.
- Tetszőleges u,v ∈ V(G) esetén van k darab éldiszjunkt u-v út, azaz k darab olyan u-v út, melyekre teljesül, hogy semelyik kettőnek nincs közös éle.

Tétel: Az alábbiak ekvivalensek:

- A G = (V,E) gráf k-szorosan összefüggő.
- Tetszőleges u,v ∈ V(G) esetén van k darab belsőleg pontdiszjunkt u-v út, azaz k darab olyan u-v út, melyekre teljesül, hogy semelyik kettőnek nincs közös pontja a kezdőponttól és a végponttól eltekintve.

Alább definiáljuk gráfok lokális él-összefüggőségét, illetve lokális összefüggőségét.

Definíció: Legyen G = (V,E) gráf, és u,v ∈ V(G) tetszőleges csúcsok. Az u és v csúcsok közötti lokális él-összefüggőség az u és v közötti éldiszjunkt utak maximális száma, melyet λ(u,v) jelöl.

Definíció: Legyen G = (V,E) gráf, és u,v ∈ V(G) tetszőleges csúcsok. Az u és v csúcsok közötti lokális összefüggőség az u és v közötti belsőleg pontdiszjunkt utak maximális száma, melyet κ(u,v) jelöl.

Könnyen ellenőrizhető, hogy egy gráf globális él-összefüggőségi száma a lokális él-összefüggőségi számok minimuma, formálisan λ(G) = min{λ(u,v) | u,v ∈ V(G)}. Hasonlóan, egy gráf globális összefüggőségi száma megegyezik a lokális összefüggőségi számok minimumával, azaz κ(G) = min{κ(u,v) | u,v ∈ V(G)}.

Ezek után feltehetjük a kérdés, hogy ha egy hálózatban a csomópontok között nem egységes összefüggőségi feltételeket szeretnénk megkövetelni, akkor hogyan soroljuk a csomópontokat osztályokba úgy, hogy az egyes osztályokra előírt összefüggőségi követelmények már egységesek legyenek.

3. GRÁFCSOMÓPONTOK OSZTÁLYOZÁSA

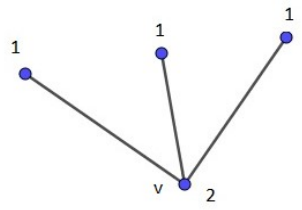
Mivel egy kommunikációs hálózatban az egyes csomópontok kiesése nem minden esetben jár ugyanakkora veszteséggel, valójában mi a csomópontok közötti lokális összefüggőséget szeretnénk vizsgálni a globális helyett. Ennek előzményeképp a csomópontokat több kockázati osztályba sorolhatjuk a hozzájuk tartozó biztonsági kockázat mértéke szerint, valamint különböző összefüggőségi követelményeket támaszthatunk az egyes osztályok elemei között. A biztonsági kockázat formalizálására a MeH ITB 12. számú ajánlása szerinti definíciót fogjuk használni gráfelméleti megfogalmazással. Először a gráf minden csúcsához definiálunk egy kezdeti kockázati értéket az alábbiak szerint. Legyen r_0 : V(G) → ℝ+ a következő függvény.

$$r_0(v) = \sum_{t \in T(v)} p(t)d(t)$$

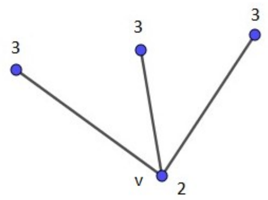
ahol $r_0(v)$ a v csúcshoz tartozó biztonsági kockázat mértéke, T(v) a v csúcsot fenyegető lehetséges támadások halmaza, p(t) a t fenyegetés bekövetkezésének valószínűsége, d(t) pedig a t fenyegetés bekövetkeztével okozott kár mértéke. Ez a kezdeti kockázat önmagában nem modellezi megfelelően a gráfcsomópontok kockázati értékét, ugyanis ha az u és v csúcsokra $r_0(u) = r_0(v)$, de az u és v szomszédjaihoz rendelt kockázati érték eltér, akkor u és v kockázati értéke is eltérhet.

Ezt a megfigyelést az alábbi ábrán szemléltetjük.

Az alábbi ábrákon a v csúcs kockázati értéke 2, azonban a 3. ábrán a v csúcs törlésével magasabb kockázati értékű csúcsok válnak elérhetetlenné, mint a 2. ábrán. Ez indokolja, hogy a v csúcs kockázati értékének meghatározásakor vegyük figyelembe N(v) elemeinek a kockázati értékét..



2. ábra



3. ábra

Legyen most r : V(G) → ℝ+ a következő függvény.

$$r(v) = r_0(v) + f(v)$$

ahol f(v) a v csúcs szomszédjainak átlagos kezdeti kockázati értéke, azaz

$$f(v) = \frac{1}{d(v)} \sum_{u \in N(v)} d(u)$$

Ezek után definiálhatjuk az u és v csúcsok távolságát, mint a D(u,v) = |r(u) – r(v)| eltérést. Bizonyítható, hogy ezzel a definícióval a D : V(G) × V(G) → ℝ függvény valóban távolság, ugyanis D nemnegatív, szimmetrikus, és teljesül rá a háromszög egyenlőtlenség. Formálisan:

- $D(u,v) = |r(u) - r(v)| \geq 0$
- $D(u,v) = |r(u) - r(v)| = |-(r(v) - r(u))| = |r(v) - r(u)| = D(v,u)$
- $D(u,v) = |r(u) - r(v)| = |r(u) - r(w) + r(w) - r(v)| \leq |r(u) - r(w)| + |r(w) - r(v)| = D(u,w) + D(w,v)$

bármely u, v csúcspárra. Amennyiben C, C'$\subseteq$ V tetszőleges csúcshalmazok, a C és C' halmazok távolságát a D(C,C') = min{D(u,v) | u∈C, v∈C'} értékkel definiáljuk.

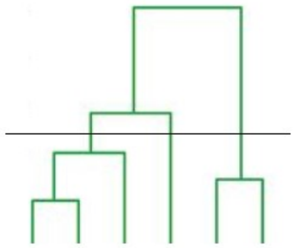
Az osztályozás problémája egy jól ismert feladatkör az adatbányászatban (F. Murtagh, 1983). Jelen esetben a gráf csomópontjait k darab $\Gamma=(C_1,\ldots,C_k)$ kockázati osztályba szeretnénk sorolni oly módon, hogy $C_1\cup\ldots\cup C_k = V$ legyen, a C_i halmazok páronként diszjunktak, és az egy osztályon belüli maximális eltérés a lehető legkisebb legyen,

azaz minimalizálni szeretnénk az M(Γ)=max{D(u,v) | u, v ∈ C} értéket.

Erről a feladatról ismert, hogy NP-teljes (T. Gonzalez, 1985), így sajnos nem várhatunk rá matematikailag hatékony algoritmust, azonban ismerünk olyan algoritmusokat, melyek a gyakorlatban jól teljesítenek. Egy egyszerű megoldást szolgáltat az agglomeratív algoritmus (F. Murtagh, 1983):

Legyen kezdetben $C_i=\{v_i\}$ minden i-re, és tegyük fel, hogy összesen k darab osztályt szeretnénk létrehozni. Mindaddig, amíg |Γ|>k, vonjuk össze azt a C_i és C_j osztályt, melyre $D(C_i,C_j)$ minimális, tehát legyen (C,C') = argmin{$D(C_i,C_j)$ | $C_i,C_j \in \Gamma$}.

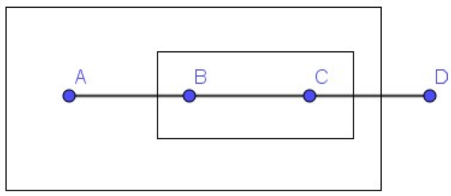
Belátható, hogy az algoritmus polinomiális, $O(n^2)$ futásidővel. A módszer előnye, hogy egyszeri futtatás után felépít egy bináris fát, melyből tetszőleges klaszterszámhoz kiolvasható a klaszterezés.



3. ábra

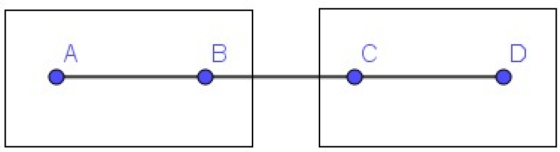
Sajnos a fenti algoritmus nem feltétlenül ad optimális megoldást.

Tekintsünk ugyanis egy gráfot a V={A,B,C,D} csúcshalmazon, melyre r(A)=1-ε, r(B)=2, r(C)=3 és r(D)=4+ε teljesül valamely ε>0 valós számra. Tegyük fel továbbá, hogy 2 kockázati osztályt szeretnénk létrehozni. Világos, hogy ekkor az algoritmus első lépésben összevonja a B, C csúcsokat, majd ezekhez hozzáveszi az A ,vagy D csúcsok valamelyikét (5. ábra),



5. ábra

Így az output a Γ={ {A,B,C},{D} } osztályozás lesz, melyben az osztályok közötti maximális távolság 2+ε. Azonban az optimális osztályozás Γ={ {A,B} },{C,D} } lenne, ekkor ugyanis a maximális távolság mindössze 1+ε (6. ábra).



6. ábra

Tehát az algoritmus által adott outputban a maximális klaszter átmérő nagyjából kétszerese az optimumnak.

6. ÖSSZEFOGLALÁS

Ebben a dolgozatban javaslatot tettünk gráfcsomópontok kockázati osztályokba sorolására olyan módon, hogy az egy osztályon belüli távolság minél kisebb legyen. Meghatároztunk egy kockázat alapú távolságfüggvényt tetszőleges gráf csomópontjai között, melyre az adatbányászatból ismert klaszterező eljárások használhatók. Ezek közül mi az agglomeratív klaszterezést ismertettük.

KÖSZÖNETNYILVÁNÍTÁS

A cikk kutatásaihoz az Új Széchenyi Terv keretein belül az EFOP-3.6.2-16-2017-00016 számú projekt biztosított forrást. A kutatás az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

ACKNOWLEDGEMENT

The research presented in this paper was carried out as part of the EFOP-3.6.2-16-2017-00016 project in the framework of the New Széchenyi Plan. The completion of this project is funded by the European Union and co-financed by the European Social Fund.

IRODALOMJEGYZÉK

T. Gonzalez, (1985), *Clustering to minimize the maximum intercluster distance*. Theoretical Computer Science, 38:293–306.

F. Murtagh, (1983), *A survey of recent advances in hierarchical clustering algorithms*. Computer Journal, 26, 4, 354-359.

S. I. Boucetta, Zs. Cs. Johanyák, L. Pokorádi, (2017), *Survey on software defined VANETs*, GRADUS 4:(1) pp. 272-283.

Jordán T., Recski A., Szeszlér D. (2004), *Rendszeroptimalizálás*, Typotex kiadó, Budapest,

Katona Y. Gy, Recski A., Szsbó Cs. (2002) *A számítástudomány alapjai*, Typotex kiadó, Budapest,

Lovász L., Pelikán J., Vesztergombi K. (2010) *Diszkrét matematika*. Typotex kiadó, Budapest

Miniszterelnöki Hivatal Informatikai Tárcaközi Bizottsága (MeH ITB) 12. számú ajánlása – Bodalki Á., Csernay A., Mátyás P., Muha L., Papp Gy., Vadász D. (1996), *Informatikai Rendszerek Biztonsági Követelményei* – Budapest.

Zentai D. (2017), *Gráfelméleti módszerek a kritikus infrastruktúra védelemben*, Hadmérnök, XII. Évfolyam 2. pp. 341-347.