# A Szemerédi lemma és alkalmazásai

Tóth Ágnes, `tothagi@cs.bme.hu`

BME, Számítástudományi és Információelméleti Tanszék

2009. április

# Szemerédi-tétel

**Szemerédi-tétel:** Tetszőleges $k > 2$ és $\varepsilon > 0$ esetén létezik $n_0$, úgy, hogy ha $n \geq n_0$ és $A \subseteq \{1, 2, \dots n\}$, $|A| > \varepsilon n$, akkor $A$ tartalmaz egy $k$ hosszú számtani sorozatot.

# Szemerédi lemma

**Jelölés:** $G = (V, E)$ gráf, $A, B \subseteq V$, $A \cap B = \emptyset$;
$d(A, B) = \frac{|E(A,B)|}{|A| \cdot |B|}$ az $A$ és $B$ közötti élsűrűség

**Definíció:** Legyen $G = (V, E)$ gráf, $A, B \subseteq V$, $A \cap B = \emptyset$, $\varepsilon > 0$;
*az* $(A, B)$ *pár* $\varepsilon$*-reguláris*, ha bármely $X \subseteq A$, $Y \subseteq B$, $|X| > \varepsilon|A|$, $|Y| > \varepsilon|B|$ esetén $|d(X, Y) - d(A, B)| < \varepsilon$.

**Definíció:** $V$ egy $\{V_0, V_1, V_2, \ldots V_k\}$ partíciója *a* $G$ *gráf* $\varepsilon$*-reguláris partíciója*, ha
$|V_0| < \varepsilon|V|$, $|V_1| = |V_2| = \ldots = |V_k|$ és
a $(V_i, V_j)$ $(i \neq j)$ párok legfeljebb $\varepsilon k^2$ kivételével $\varepsilon$-regulárisak.

# Szemerédi lemma

**Jelölés:** $G = (V, E)$ gráf, $A, B \subseteq V$, $A \cap B = \emptyset$;
$d(A, B) = \frac{|E(A,B)|}{|A| \cdot |B|}$ az $A$ és $B$ közötti élsűrűség

**Definíció:** Legyen $G = (V, E)$ gráf, $A, B \subseteq V$, $A \cap B = \emptyset$, $\varepsilon > 0$;
*az* $(A, B)$ *pár* $\varepsilon$*-reguláris*, ha bármely $X \subseteq A$, $Y \subseteq B$, $|X| > \varepsilon|A|$,
$|Y| > \varepsilon|B|$ esetén $|d(X, Y) - d(A, B)| < \varepsilon$.

**Definíció:** $V$ egy $\{V_0, V_1, V_2, \ldots V_k\}$ partíciója *a* $G$ *gráf* $\varepsilon$*-reguláris partíciója*, ha
$|V_0| < \varepsilon|V|$, $|V_1| = |V_2| = \ldots = |V_k|$ és
a $(V_i, V_j)$ $(i \neq j)$ párok legfeljebb $\varepsilon k^2$ kivételével $\varepsilon$-regulárisak.

**Szemerédi Regularitási Lemma:** Tetszőleges $\varepsilon > 0$ valós és $m \geq 1$ egész számhoz létezik $M > 0$ egész úgy, hogy bármely $G = (V, E)$ gráfhoz, amire $|V(G)| > m$, létezik $\{V_0, V_1, V_2, \ldots V_k\}$ $\varepsilon$-reguláris partíciója, melyre $m \leq k \leq M$.

A bizonyítás ötlete.

$V(G)$ partícióihoz definiálunk egy $q$ indexet, melyre a következők teljesülnek:
- $0 \leq q \leq 1$;
- ha a $\mathcal{P}'$ partíció finomítása a $\mathcal{P}$ partíciónak, akkor $q(\mathcal{P}') \geq q(\mathcal{P})$;
- ha a $\mathcal{P}$ partíció nem $\varepsilon$-reguláris, akkor van olyan $\mathcal{P}'$ finomítása, hogy

$q(\mathcal{P}') - q(\mathcal{P}) \geq \frac{\varepsilon^5}{2}$,
továbbá ha $\mathcal{P}$-ben $k$ osztály volt, akkor $\mathcal{P}'$-ben $l$ osztály lesz, ahol
$k \leq l \leq k4^k$
(és a kivételes halmaz sem nő meg nagyon: $|V_0'| \leq |V_0| + \frac{|V(G)|}{2^k}$).

Kiindulunk egy partícióból és ezt addig finomítjuk, míg a végén egy $\varepsilon$-reguláris partíciót kapunk. $q$ tulajdonságaiból adódóan ehhez véges sok lépés elég, a végén kapott partícióosztályok száma $\varepsilon$ és $m$ függvényével felülről becsülhető.

# Kulcslemma

**Észrevétel:** Legyen $(A, B)$ $\varepsilon$-reguláris pár $d = d(A, B)$ sűrűséggel, $Y \subseteq B$, $|Y| \geq \varepsilon|B|$; ekkor $A$-ban legfeljebb $\varepsilon|A|$ csúcs kivételével bármely csúcsnak legalább $(d - \varepsilon)|Y|$ szomszédja van $Y$-ban.

# Kulcslemma

**Észrevétel:** Legyen $(A, B)$ $\varepsilon$-reguláris pár $d = d(A, B)$ sűrűséggel, $Y \subseteq B$, $|Y| \geq \varepsilon|B|$; ekkor $A$-ban legfeljebb $\varepsilon|A|$ csúcs kivételével bármely csúcsnak legalább $(d - \varepsilon)|Y|$ szomszédja van $Y$-ban.

**Definíció:** Legyen $G = (V, E)$ gráf, $\{V_0, V_1, V_2, \ldots V_k\}$ a $G$ egy $\varepsilon$-reguláris partíciója, $|V_i| = l$ $(1 \leq i \leq k)$, továbbá $0 < d \leq 1$;
tekintsük a következő $R$ gráfot:
$V(R) = \{V_1, V_2, \ldots V_k\}$,
$E(R) = \{\{V_i, V_j\} : (V_i, V_j) \text{ } \varepsilon\text{-reguláris, } d \text{ sűrűséggel}\}$
ekkor az $R$ gráfot a $G$ gráf $(\varepsilon, l, d)$ paraméterekhez és az adott partícióhoz tartozó *redukált gráfjának* nevezzük.

**Definíció:** Adott $s$ pozitív egész és $R$ gráf;
*$R^s$ az $R$ gráf $s$-szeres felfújtja* az a gráf, melyre
$V(R^s) = \{r_i^j : r_i \in V(R), 1 \leq j \leq s\}$,
$E(R^s) = \{\{r_i^j, r_k^l\} : \{r_i, r_k\} \in E(R)\}$.

**Kulcslemma:** Bármely $0 < d \leq 1$ valós és $\Delta \geq 1$ egészhez létezik $\varepsilon_0 > 0$ a következő tulajdonsággal:
ha $G$ tetszőleges gráf, $H$ egy olyan gráf, melyre $\Delta(H) \leq \Delta$, $s$ pozitív egész, az $R$ pedig a $G$ gráf redukált gráfja $(\varepsilon < \varepsilon_0, l \geq \frac{s}{\varepsilon_0}, d)$ paraméterekkel, akkor $H \subseteq R^s$ esetén $H \subseteq G$ is teljesül.

**Kulcslemma:** Bármely $0 < d \leq 1$ valós és $\Delta \geq 1$ egészhez létezik $\varepsilon_0 > 0$ a következő tulajdonsággal:
ha $G$ tetszőleges gráf, $H$ egy olyan gráf, melyre $\Delta(H) \leq \Delta$, $s$ pozitív egész, az $R$ pedig a $G$ gráf redukált gráfja ($\varepsilon < \varepsilon_0, l \geq \frac{s}{\varepsilon_0}, d$) paraméterekkel, akkor $H \subseteq R^s$ esetén $H \subseteq G$ is teljesül.

Bizonyítás vázlat.

- jön $d$ és $\Delta$; $\varepsilon_0$ legyen olyan, hogy $\varepsilon_0 < d$ és $\frac{\Delta+1}{(d-\varepsilon_0)^\Delta}\varepsilon_0 \leq 1$;

- $G$, $H$, $s$ és $R$ az állítás szerinti ($\{V_0, V_1, V_2, \dots V_k\}$ a $G$ megfelelő $\varepsilon$-reguláris partíciója), $H \subseteq R^s$;

- legyen $V(H) = \{u_1, u_2, \dots u_h\}$, célunk, hogy minden egyes $u_i$-ra ($0 \leq i \leq h$) megadjunk egy $v_i \in V(G)$ csúcsot, úgy, hogy $\{u_i, u_j\} \in E(H)$ esetén $\{v_i, v_j\} \in E(G)$ teljesüljön;

- megyünk végig $H$ csúcsain, és folyamatosan rendeljük őket az egyes $v_i$ csúcsokhoz, közben minden $u_i$-re nyilvántartunk egy $Y_i$ (folyamatosan szűkülő) halmazt, mely $v_i$ lehetséges értékeit tartalmazza, kezdetben $Y_i^0$ azon $V_j$, melynek felfújtjában $u_i$ benne van;

- a csúcsok hozzárendelésénél figyelnünk kell, hogy az $Y_i$ halmazok elegendőek nagyok maradjanak a későbbi csúcsok rögzítéséhez;

- ehhez elegendő, ha $v_i$ megválasztásakor $|Y_i| \geq \Delta\varepsilon l + s$, továbbá a későbbiekre $|Y_i| \geq \varepsilon l$, ehhez pedig elég ha $(d-\varepsilon)^\Delta l - \Delta\varepsilon l \geq s$,

- $\varepsilon_0$ fenti megválasztásával ez teljesül.

# Alkalmazások: Erdős-Stone tétel

**Turán tétel:** Ha egy $n$ csúcsú $G$ gráf nem tartalmaz $K_r$-et, akkor $|E(G)| \leq |E(T_{r-1}(n))|$. Ha pedig $|E(G)| = |E(T_{r-1}(n))|$, akkor $G \simeq T_{r-1}(n)$. (Ahol $T_{r-1}(n)$ az $n$ csúcsú, $r-1$ osztályú Turán-gráfot jelöli.)

# Alkalmazások: Erdős-Stone tétel

**Turán tétel:** Ha egy $n$ csúcsú $G$ gráf nem tartalmaz $K_r$-et, akkor $|E(G)| \leq |E(T_{r-1}(n))|$. Ha pedig $|E(G)| = |E(T_{r-1}(n))|$, akkor $G \simeq T_{r-1}(n)$. (Ahol $T_{r-1}(n)$ az $n$ csúcsú, $r-1$ osztályú Turán-gráfot jelöli.)

**Erdős-Stone tétel:** Bármely $r \geq 2$, $s \geq 1$, $\gamma > 0$-ra létezik $n_0$, hogy minden $n \geq n_0$ csúcsú és legalább $|E(T_{r-1}(n))| + \gamma n^2$ élű gráf tartalmaz $K_r^s$ részgráfot.

Bizonyítás vázlat.

- jön $r$, $s$ és $\gamma$;

- a Kulcslemma $d = \gamma$ és $\Delta = \Delta(K_r^s)$-hez ad egy $\varepsilon_0$-t;

- továbbá legyen $m$ olyan, hogy $m > \frac{1}{\gamma}$,
$\varepsilon$ pedig olyan, hogy $0 < \varepsilon \leq \varepsilon_0$ és $2\gamma - \varepsilon^2 - 4\varepsilon - d - \frac{1}{m} > 0$;

- Szemerédi Lemma $\varepsilon$-hoz és $m$-hez ad $M$-et;

- $n$ legyen olyan, hogy $n \geq n_0 = \frac{Ms}{\varepsilon_0(1-\varepsilon)}$;

- jön $G$, egy $G$ gráfhoz, melyre $|V(G)| = n$ a kapott $k(+1)$ elemű $\varepsilon$-partícióban az osztályok $l$ méretére ($n$ választása miatt) teljesül, hogy $l \geq \frac{s}{\varepsilon_0}$, így a Kulcslemma alkalmazható;

- tehát ha $K_r \subseteq R$, akkor $K_r^s \subseteq R^s$ és a Kulcslemma miatt $K_r^s \subseteq G$ ($\varepsilon_0$ jól lett megválasztva);

- így elég belátni, hogy $K_r \subseteq R$;

- $|E(G)| \leq \frac{1}{2}(\varepsilon n)^2$ ($V_0$-on belüli élek)
$+\varepsilon nkl$ ($V_0$ és $V_i$-k ($i \neq 0$) közötti élek)
$+\varepsilon k^2 l^2$ (nem $\varepsilon$-reguláris $V_i$,$V_j$ közötti élek)
$+\frac{1}{2}k^2 dl^2$ (kisebb, mint $d$-sűrűségű $\varepsilon$-reguláris $V_i$,$V_j$ közötti élek)
$+|E(R)|l^2$ (legalább $d$-sűrűségű $\varepsilon$-reguláris $V_i$,$V_j$ közötti élek)
$+\frac{1}{2}l^2 k$ ($V_i$-ken ($i \neq 0$) belüli élek)

- ebből elég nagy $n$-re: $|E(R)| \geq \frac{1}{2}k^2 \frac{r-2}{r-1} \geq |E(T_{r-1}(k))|$, így $K_r \subseteq R$.

# Alkalmazások: Chvátal-Rödl-Szemerédi-Trotter tétel

**Ramsey tétel:** Bármely $k$-ra létezik egy olyan legkisebb $R(k)$, hogy $n \geq R(k)$ esetén $K_n$ éleit két színnel színezve biztosan lesz egyszínű $K_k$.

**Tétel (Erdős, illetve Erdős-Szekeres):**
Ha $k \geq 3$, akkor $2^{k/2} \leq R(K) \leq 4^k$.

**Definíció:** $R(H)$ az a legkisebb pozitív egész szám, amire igaz, hogy $n \geq R(H)$ esetén $K_n$ éleit két színnel színezve biztosan lesz egyszínű $H$ részgráf (nem kell, hogy az adott színű részgráfnak feszített részgráfja legyen).

# Alkalmazások: Chvátal-Rödl-Szemerédi-Trotter tétel

**Ramsey tétel:** Bármely $k$-ra létezik egy olyan legkisebb $R(k)$, hogy $n \geq R(k)$ esetén $K_n$ éleit két színnel színezve biztosan lesz egyszínű $K_k$.

**Tétel (Erdős, illetve Erdős-Szekeres):**
Ha $k \geq 3$, akkor $2^{k/2} \leq R(K) \leq 4^k$.

**Definíció:** $R(H)$ az a legkisebb pozitív egész szám, amire igaz, hogy $n \geq R(H)$ esetén $K_n$ éleit két színnel színezve biztosan lesz egyszínű $H$ részgráf (nem kell, hogy az adott színű részgráfnak feszített részgráfja legyen).

**Tétel (Chvátal, Rödl, Szemerédi, Trotter):** Minden pozitív egész $\Delta$-hoz létezik $c$ valós konstans, hogy tetszőleges $H$ gráfra, melyre $\Delta(H) \leq \Delta$ igaz, hogy $R(H) \leq c|V(H)|$.

# Algoritmikus kérdések

**Tétel (Alon, Duke, Lefmann, Rödl, Yuster):**
A következő eldöntési probléma co-NP-teljes:
input: $G$ gráf, $\varepsilon > 0$, $k \geq 1$ és a csúcsok egy partíciója $k + 1$ részre;
kérdés: a megadott partíció $\varepsilon$-reguláris-e?

# Algoritmikus kérdések

**Tétel (Alon, Duke, Lefmann, Rödl, Yuster):**
A következő eldöntési probléma co-NP-teljes:
input: $G$ gráf, $\varepsilon > 0$, $k \geq 1$ és a csúcsok egy partíciója $k + 1$ részre;
kérdés: a megadott partíció $\varepsilon$-reguláris-e?

**Tétel (Sz, ADLRY):** Tetszőleges $\varepsilon > 0$ valós és $m \geq 1$ egész számhoz létezik $\hat{M} > 0$ egész úgy, hogy bármely $G = (V, E)$ gráfhoz, amire $n = |V(G)| > \hat{M}$, létezik $\{V_0, V_1, V_2, \dots V_k\}$ $\varepsilon$-reguláris partíció, melyre $m \leq k \leq \hat{M}$.
Tetszőleges rögzített $\varepsilon$-ra és $m$-re egy ilyen partíció megtalalálható $O(M(n))$ időben, ahol $M(n)$ két $n \times n$-es 0, 1 elemű mátrix összeszorzásának ideje az egészek felett, $M(n) = O(n^{2.376})$.
(Polinom sok párhuzamos processzort használva $O(\log n)$ idő elég EREW PRAM[1]-okkal.)

[1] Exclusive Read Exclusive Write Parallel Random Access Machine

# Algoritmikus alkalmazások

**Tétel (ADLRY):** Bármely $\delta > 0$-ra létezik $c > 0$ úgy, hogy tetszőleges $m$-re, és $m$ élű $H$ gráfra igaz, hogy ha egy $G$ gráfnak $n \geq cm$ csúcsa és legalább $\delta n^2$ éle van, akkor $G$ tartalmaz egy $H$ -val topologikusan izomorf részgráfot, melyben $H$ minden éle $4$ hosszú úttal van helyettesítve. $G$-ben ilyen részgráf polinom időben megtalálható.

# Algoritmikus alkalmazások

**Tétel (ADLRY):** Bármely $\delta > 0$-ra létezik $c > 0$ úgy, hogy tetszőleges $m$-re, és $m$ élű $H$ gráfra igaz, hogy ha egy $G$ gráfnak $n \geq cm$ csúcsa és legalább $\delta n^2$ éle van, akkor $G$ tartalmaz egy $H$ -val topologikusan izomorf részgráfot, melyben $H$ minden éle $4$ hosszú úttal van helyettesítve. $G$-ben ilyen részgráf polinom időben megtalálható.

**Tétel (CRSzT, ADLRY):** Minden pozitív egész $\Delta$-hoz létezik $c$ valós konstans úgy, hogy tetszőleges $H$ gráfra, melyre $\Delta(H) \leq \Delta$ és tetszőleges $G$ gráfra, melynek legalább $c|V(H)|$ csúcsa van teljesül, hogy $G$ vagy $G$ komplementere tartalmaz $H$-val izomorf részgráfot. Egy ilyen részgráf polinom időben megtalálható.

# Alkalmazások klaszterezésben és képfeldolgozásban

**Heurisztikus algoritmus a klaszterezésre: (Sperotto, Pelillo)**

● adott egy súlyozott élű $G$ gráf, csúcsai az egyedek, élek súlya a hasonlóság mértéke;

● elkészítjük a $G$ gráf egy $\varepsilon$-reguláris partícióját - legalábbis annak valamilyen közelítését, nem túl sok partícióval (egy lépésben nem osztunk túl sok részre, továbbá ha egy előre definiált határt meghaladna a partíciók száma, akkor leállunk);

● majd a partícióhoz tartozó $R$ redukált gráffal dolgozunk tovább (az egyes partíciókat még tovább daraboljuk a partíción belüli élek alapján);

● az $R$ gráfra alkalmazunk valamilyen klaszterezési algoritmust (pl. Pavan, Pelillo domináns halmaz ötletét);

● a $R$ osztályozása alapján klaszterezzük a $G$ gráfot (az eldobott kivételes halmaz elemeit a kialakuló csoportok valamelyikéhez illesztjük).

Tapasztalat:

különböző adatbázisokon elvégzett mérések alapján az algoritmus hasonlóan jó minőségű klaszterezést épített fel, mint az eredeti gráfokon futtatott klaszterező, viszont sokkal ($\mathbf{4}$-$\mathbf{20}$-szor) gyorsabban.

## Felhasznált irodalom:

J. Komlós, A. Shokoufandeh, M. Simonovits, E. Szemerédi: Szemerédi's regularity lemma and its applications in graph theory,
Combinatorics, Paul Erdös is eighty, Vol. 2 (Keszthely, 1993), 295–352, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.
http://www.renyi.hu/~miki/komsimshoszem.pdf

J. Komlós, M. Simonovits: Szemerédi's regularity lemma and its applications in graph theory, http://www.renyi.hu/~miki/komsimw.pdf

M. Simonovits: Irodalom a Szemerédi lemmához
http://www.renyi.hu/~miki/Irodalom.html

G. Simonyi: Gráfok és hipergráfok előadásjegyzet

N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster: The algorithmic aspects of the Regularity Lemma, Proc. 33 IEEE FOCS, Pittsburgh, IEEE (1992), 473-481. Also: J. of Algorithms 16 (1994), 80-109.
http://www.math.tau.ac.il/~nogaa/PDFS/reg5.pdf

A. Sperotto, M. Pelillo: Szemerédi's Regularity Lemma and Its Applications to Pairwise Clustering and Segmentation, Chapter in: Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer Berlin, 2007.

http://videolectures.net/gbr07_pelillo_srl/

M. Pavan and M. Pelillo: Dominant sets and pairwise clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(1):167-172, 2007.

http://www.dsi.unive.it/˜pelillo/papers/PAMI%202007.pdf

# Kiegészítés

A $(6, 3)$-probléma:

Tétel (Ruzsa, Szemerédi): Ha $H_n$ egy $3$-uniform hipergráf $n$ csúcson úgy, hogy nincs olyan $3$ éle, melyek uniója legfeljebb $6$ pontot adna, akkor $|E(H_n) = o(n^2)|$.

# Kiegészítés

A $(6,3)$-probléma:

Tétel (Ruzsa, Szemerédi): Ha $H_n$ egy $3$-uniform hipergráf $n$ csúcson úgy, hogy nincs olyan $3$ éle, melyek uniója legfeljebb $6$ pontot adna, akkor $|E(H_n) = o(n^2)|$.

Erdős-Sós sejtés: Tetszőleges $n$ csúcsú és több, mint $(k-1)n/2$ élű gráf részgráfként tartalmaz minden $k$ élű fát.

Ajtai, Komlós, Simonovits, Szemerédi: Az Erdős-Sós sejtés igaz nagy $n$-ekre.

Hipergráf regularitási lemma:
W.T. Gowers: Hypergraph regularity and the multidimensional Szemerédi theorem,
Annals of mathematics, Vol. 166, No. 3, 2007, page 897-946.
http://www.dpmms.cam.ac.uk/˜wtg10/hypersimple4.pdf

G. Elek, B. Szegedy: A measure-theoretic approach to the theory of dense hypergraphs,
http://arxiv.org/pdf/0810.4062v2

Hipergráf regularitási lemma:
W.T. Gowers: Hypergraph regularity and the multidimensional Szemerédi theorem,
Annals of mathematics, Vol. 166, No. 3, 2007, page 897-946.
http://www.dpmms.cam.ac.uk/˜wtg10/hypersimple4.pdf

G. Elek, B. Szegedy: A measure-theoretic approach to the theory of dense hypergraphs,
http://arxiv.org/pdf/0810.4062v2

Regularitási lemma ritka gráfokra ($|E(G)| > cn^{2-\alpha}$):
Y. Kohayakawa: Szemerédi's regularity lemma for sparse graphs,
Foundations of Computational Mathematics (Berlin, Heidelberg),
Springer-Verlag, 1997, pp. 216-230
http://www.ime.usp.br/˜yoshi/MSs/FoCM/sparse.ps.gz

Y. Kohayakawa V. Rödl: Regular pairs in sparse random graphs,
Random Structures & Algorithms, 2003, vol. 22, no. 4, pg. 359-434
http://www.ime.usp.br/˜yoshi/MSs/BigI/rpI.ps.gz