Není hvězda jako hvězda: identifikace klíčových aktérů v sociálních sítích[1]

No Star Like the Other: The Measurement of Centrality in Social Networks

Tomáš Diviák

ABSTRACT The concept of centrality and centrality measures are well-known and frequently used in social network analysis. They are also implemented in numerous software packages. However, that does not mean that it is easy to apply them correctly. This paper aims to introduce the most frequently used centrality measures, but more importantly to point out the problems related to their application and to sketch potential solutions for these problems. First, three basic centrality measures are introduced: degree, betweenness, and closeness. There are three broad categories of issues with centrality measures. These categories are: inadequate operationalisation of centrality measures, explanation of their distribution, and interdependence of observation in statistical modelling. A typology of flows in the network is presented as a potential solution allowing for transparent operationalisation. The so-called positional approach is another potential solution allowing for conceptually and computationally rigorous definition of centrality measures. Lastly, statistical models for network data are introduced as a way to deal with interdependence of observations. In the conclusion, challenges for measuring centrality in bipartite and multiplex networks are discussed.

KEYWORDS social network analysis, centrality measures, methodology

Úvod

Analýza sociálních sítí (dále SNA[2]) získává v českém a slovenském (nejen) sociálněvědním prostředí na popularitě. V posledních letech dochází k pravidelnému setkávání výzkumníků a výzkumnic zabývajících se tímto přístupem během kolokvií a workshopů, a nejen to – tito lidé také s narůstající intenzitou využívají SNA ve svém výzkumu. V domácích i mezinárodních odborných periodikách tak vycházejí články zabývající se aplikací SNA ve výzkumu bezdomovectví (Daňková, Bernard a Vašát 2019; Vašát a Bernard 2015),

Sociální studia / Social Studies 1/2021. S. 101–120. ISSN 1214-813X.

1 Tento výzkum byl podpořen Grantovou agenturou ČR (Grant 17-22978Y). Autor tohoto textu děkuje Renátě Topinkové a oběma anonymním recenzentům za nesmírně cennou a detailní zpětnou vazbu.

2 Z anglického *social network analysis*.

neziskového sektoru (Císař a Navrátil 2015; Mazák a Diviák 2018; Ocelík et al. 2019), organizovaného zločinu (Diviák, Dijkstra a Snijders 2018, 2019), ale také třeba funkční konektivity lidského mozku (Klobušiaková, Mareček, Fousek, Výtvarová a Rektorová 2019) a objevují se i statě metodologické povahy (Diviák 2017).

K posledně jmenovaným, tedy k metodologickým studiím, přispívá i tento text, jehož cílem je představit koncept centrality a jeho měření v SNA. Obecně řečeno, centralita označuje relativní význam uzlů (aktérů) v dané síti a měří se zpravidla pomocí tzv. měr centrality. Míry centrality jsou velice populární díky tomu, že jsou intuitivně snadno uchopitelné, a také proto, že mají téměř bezprostřední využití v řadě výzkumných i praktických oblastí od identifikace názorových vůdců, přes diagnostiku pracovních kolektivů až po narušování kriminálních sítí. Intuitivní uchopitelnost a praktické využití však bohužel neznamenají, že je měření centrality prosté jakýchkoliv problémů. Ačkoliv je problematika měření centrality součástí SNA od jejích prvopočátků a je součástí snad každého úvodního textu, vývoj v této oblasti jde stále kupředu a nové poznatky se do učebních textů již v zásadě nedostávají.

Míry centrality jsou také široce a snadno dostupné ve většině softwarových balíčků určených pro SNA. To se sice může jevit jako jejich nesporná výhoda, má to však i svou stinnou stránku. To, že lze míry centrality spočítat pomocí jednoduché funkce, či dokonce prostřednictvím několika málo kliknutí, může totiž uživatele svádět ke zkratkovitému či nepřesnému užívání těchto nástrojů. V tomto ohledu SNA zapadá do širší problematiky nevhodného využívání a potažmo i zneužívání formálních metod v sociálních vědách. V běžném kvantitativním výzkumu se v posledních letech hovoří o problematice špatného užívání a zneužívání p-hodnot pro statistickou inferenci (srov. Benjamin et al. 2017; Nuzzo 2014; v českém prostředí např. Soukup 2019) nebo o replikační krizi studií kvůli jejich nedostatečné statistické síle (srv. Camerer et al. 2018; Open Science Collaboration 2015). Relativní novost spojená s efektními (nikoliv nutně efektivními) vizualizacemi a snadným výpočtem desítek měr centrality může vést k podobným nešvarům i v SNA.

Cílem tohoto textu je na příkladech nejpoužívanějších měr centrality nejprve upozornit na největší úskalí související s jejich aplikací. Na základě nových poznatků a metodologických postupů jsou následně předestřena možná řešení těchto problémů tak, aby byly výsledky generované pomocí měr centrality teoreticky relevantní a metodologicky validní a reliabilní. Tento text tedy nemá ambici být vyčerpávající přehledovou studií ani nemá za cíl představit jednu konkrétní metodologickou inovaci. Cílem je spíše zmapovat metodologické problémy a jejich aktuální řešení a souhrnně je představit českému a slovenskému odbornému publiku. Článek má proto poněkud nezvyklou strukturu, kdy je nejprve definována základní terminologie a míry centrality, následně jejich největší problémy a posléze řešení těchto problémů. Nakonec jsou představena možná rozšíření měr centralit na specifické typy síťových dat.

Základní terminologie

Než se přesuneme k výkladu samotných měr centrality a následně k jejich problémům a potenciálním řešením, je namístě vymezit, co v této stati rozumíme sítí a jak se toto pojetí liší od jiných přístupů, třebaže využívají podobnou terminologii. Za prvé, termín síť je často

užíván v rámci některých soudobých teoretických proudů, které reprezentují úvahy o síťové společnosti (např. Castells 2009) nebo o aktérech-sítích (např. Latour 2007). V těchto úvahách sítě reprezentují jakýsi zcela nový způsob organizace sociálních vztahů, který se vyznačuje pomíjivostí, flexibilitou, horizontálností a dynamikou v protikladu k předchozím hierarchičtějším a statičtějším formám sociálního řádu.[3] Problémem takovýchto úvah je především jejich vysoká míra abstrakce, metaforičnosti a nejasnosti, která jen stěží umožňuje jejich empirické testování. Přes řadu zdánlivých styčných bodů se tak v SNA s takovýmito velkými teoriemi pracuje jen vzácně. Za druhé, v souvislosti s rozmachem sociálních médií, jako jsou Facebook či Twitter, se v médiích i na veřejnosti často mluví o sociálních sítích, přičemž míněna jsou právě sociální média. Ačkoliv lze pomocí SNA analyzovat sítě přátelství uživatelů na Facebooku nebo retweetování či sledování účtů na Twitteru, sociální sítě v SNA odkazují obecně ke vztahům a interakcím mezi lidmi, potažmo jinými sociálně žijícími živočichy.

Síť v SNA je definována jako množina uzlů a vazeb mezi nimi.[4] Uzly mohou představovat libovolné entity – města, státy, skupiny, různé živočichy, abstraktní jednotky (symboly, slova) apod., typicky však v sociálních sítích reprezentují aktéry, tj. lidské bytosti schopné mezi sebou nějakým způsobem interagovat a vytvářet vzájemné vztahy. Tyto vztahy a interakce jsou pak reprezentovány vazbami (též hranami). Vazby mají dvě zásadní vlastnosti – orientaci a sílu. Podle orientace lze vazby rozlišit na orientované (u nichž je možné rozlišit, od kterého uzlu ke kterému směřují) a neorientované (kde takové rozlišení možné není). Z hlediska síly lze vazby obecně rozlišit na vazby binární (tj. kde vazba mezi každým párem uzlů indikuje přítomnost či absenci dané relace) a vazby ohodnocené, přičemž sílu vazeb lze popsat buď ordinální stupnicí, nebo intenzitou vyjádřenou kardinálně. Například v síti spolužáků ve třídě je relace „sedět ve stejné lavici“ reprezentována vazbami binárními a neorientovanými, neboť sedí-li žák A s žákem B, pak nutně sedí i B s A. Oproti tomu relaci „pomáhá s úkoly“ zachycují vazby orientované, neboť symetrie takové relace není samozřejmá a záleží na konkrétním designu, zda síla vazeb bude binární (pomáhá/nepomáhá), nebo třeba zachycena nějakou ordinální stupnicí (v rámci jednoho předmětu – v několika předmětech – ve většině předmětů – ve všech předmětech atd.). V síti států ve světě je příkladem ohodnocené orientované relace třeba objem exportu z jednoho státu do druhého, neboť je možné určit směr a sílu vazby lze kvantifikovat například v množství peněz. Neorientovanou binární relací mezi státy pak může být například „mít spolu uzavřenou obchodní dohodu“. Výslednou síť lze zobrazit vizuálně ve formě grafu (sociogramu) nebo prostřednictvím matice sousednosti. Matice sousednosti má rozměry $n \times n$, kde n je počet uzlů, přičemž řádky i a sloupce j této matice reprezentují jednotlivé uzly a konkrétní prvky matice a_{ij} značí hodnoty či (ne)přítomnost vazby mezi uzly i a j.

---

3 Tyto síťově inklinující teorie jsou vnitřně značně diferencované a samy o sobě komplexní, proto zde uvádím jen jejich jakýsi nejmenší společný jmenovatel. Konkrétní předpoklady a teze lze dohledat v primární literatuře.

4 Definice v tomto odstavci jsou v SNA konsenzuální a lze tak najít podobnou (s drobnými odchylkami) terminologii napříč všemi úvodními texty k SNA. Blíže viz texty Borgattiho a kolegů (2013), Prell (2011), Robinse (2015) či Wassermana a Faust (1994).

V každém případě je však dobré pro začátek zachovat dvě podmínky pro definování množiny uzlů a vazeb mezi nimi. Co se uzlů týče, měly by reprezentovat entity stejného typu – tj. jen žáky a jen státy, a to tak, aby každý uzel měl hypoteticky stejnou možnost mít vazbu k ostatním. Je například značně problematické zahrnout do sítě žáků i vyučujícího, neboť u něj nemá smysl uvažovat o tom, že by mohl mít vazbu v relaci „sedět ve stejné lavici" nebo že by mohl dostávat příchozí vazby v rámci relace „pomáhá s úkoly". Podobně jako uzly, i vazby by se měly týkat jen jedné dobře definované relace a měly by mít stejně vymezenou orientaci, tj. měly by být buď všechny orientované, nebo všechny neorientované a měly by mít také jednotně definovanou sílu, tedy analogicky k orientaci by i síla měla být buď jen binární, nebo jen ordinální, nebo jen kardinální. V souladu s těmito podmínkami lze definovat unipartitní uniplexní síť, která je nejjednodušším a nejběžnějším typem sítě, v níž je pouze jeden typ uzlů a jeden typ vazeb mezi nimi. Ve zbytku tohoto textu budeme vycházet z této nejjednodušší síťové reprezentace. V závěrečné části se pak stručně podíváme i na složitější a méně obvyklé typy síťových dat.

Míry centrality

Koncept centrality lze definovat široce jako relativní význam uzlu v síti (Borgatti 2005; Borgatti, Everett a Johnson 2013; Freeman 1978; Prell 2011; Robins 2015; Wasserman a Faust 1994) – čím centrálnější daný uzel je, tím větší význam v rámci sítě má. V této souvislosti se s identifikací centrálních aktérů či s měřením centrality obecně pojí dlouhá řada substantivních výzkumných poznatků v řadě domén. Pro příklad uveďme problematiku kriminálních sítí (Morselli 2009), názorového vůdcovství (Valente a Pumpuang 2007) či veřejného zdraví (Luke a Harris 2007). V této části představím tři nejpoužívanější a konsenzuálně nejzásadnější míry centrality a stručně popíši jejich vlastnosti včetně výhod a nevýhod (viz např. Borgatti 2005; Borgatti et al. 2013; Brandes 2016; Landherr, Friedl a Heidemann 2010; Robins 2015). Jsou jimi stupeň, mezilehlost a blízkost.[5] Jedná se také o míry centrality, jež jsou pojednány v dnes již klasickém Freemanově (1979) textu k této problematice.

Stupeň

Nejsnazším a nejintuitivnějším způsobem, jak operacionalizovat koncept centrality v SNA, je prostřednictvím počtu vazeb, které daný uzel má. Čím více vazeb má, tím centrálnější daný uzel je. Toto zachycuje míra centrality zvaná stupeň (angl. *degree*), která udává počet vazeb, jež má daný uzel. Stupeň uzlu i lze spočítat na základě matice sousednosti s prvky a_{ij} následovně:

$$d_i = \sum_j a_{ij}$$

5 Existuje ještě celá řada dalších a dalších měr centrality, které jsou přehledně pojednány v úvodních textech k SNA (např. Borgatti et al. 2013; Prell 2011; Wasserman a Faust 1994). V těchto textech je k dispozici také bližší didaktický rozbor a instrukce k výpočtům těchto měr. Zde se soustředím především na základní konceptuální výklad.

V případě, že jsou vazby v síti orientované, stupeň lze rozdělit na vstupní (*indegree*) a výstupní (*outdegree*), přičemž jak již názvy napovídají, vstupní stupeň zachycuje počet vazeb příchozích – orientovaných směrem k danému uzlu –, kdežto výstupní stupeň zachycuje počet vazeb vyslaných – orientovaných směrem od daného uzlu. Jsou-li v matici sousednosti vazby odchozí zachyceny v řádcích a vazby příchozí zachyceny ve sloupcích, pak je vstupní stupeň uzlu *i* sloupcovým součtem a analogicky pro řádkový součet v případě stupně výstupního. V případě ohodnocených vazeb pak lze jako ohodnocený stupeň (*weighted degree*) chápat součet síly vazeb daného uzlu.

Stupeň je základní měrou centrality a má řadu konceptuálních výhod – je snadný na vypočtení, porozumění i komunikaci výsledků. Stupeň má navíc velice intuitivní interpretaci na úrovni aktérů, kdy jej lze jednoduše interpretovat z perspektivy daného aktéra, např. uzel s vysokým stupněm v síti žáků s vazbami reprezentující přátelství je vysoce sociabilní a oblíbený aktér. V případě orientovaných vazeb pak lze rozlišit uzly aktivní (vysoký výstupní stupeň) a populární (vysoký vstupní stupeň). Nevýhodou stupně jako míry centrality je fakt, že zohledňuje pouze přímá spojení uzlů a nikterak nebere v potaz strukturu sítě či centralitu partnerů. Takto může například z analýzy vyjít jako centrální uzel aktér s velkým počtem vazeb v podskupině, která je jinak izolovaná od zbytku sítě, což celkový význam tohoto uzlu značně omezuje. Širší strukturu sítě však nějakým způsobem zohledňují následující míry centrality.

Mezilehlost

V pojetí mezilehlosti (*betweenness*) je centrální takový uzel, přes nějž proudí relativně velké množství toků informací či zdrojů v síti. Mezilehlost určuje centrální uzly na základě geodetických vzdáleností. Geodetická vzdálenost je taková neopakující se sekvence vazeb a uzlů (v teorii grafů cesta) mezi dvojicí uzlů, která ze všech možných takových sekvencí (cest) mezi touto dvojicí zahrnuje nejméně vazeb. Jednoduše řečeno, geodetická vzdálenost je nejkratší cesta, která spojuje danou dvojici uzlů. Vysokou mezilehlost má pak takový uzel, přes nějž prochází mnoho geodetických cest, tedy uzel přemosťující spojení mezi řadou dalších uzlů v síti. Formálně lze tak mezilehlost uzlu *j* vyjádřit jako:

$$b_j = \sum_{i \neq k} \frac{g_{ijk}}{g_{ik}}$$

Kde g_{ijk} značí geodetické cesty mezi *i* a *k* procházející přes *j* a g_{ik} pak všechny pozorované cesty mezi *i* a *k*. Jednoduše řečeno, jedná se o poměr mezi přítomnými geodetickými cestami mezi všemi možnými páry aktérů (vyjma uzlu *j*), které procházejí přes uzel *j*, ku počtu všech takových cest včetně těch, které uzel *j* nezahrnují.[6] Mezilehlost tak dokáže identifikovat jako centrálního takového aktéra, který sice nemusí být vůbec výjimečně aktivní (stačí

6 Pokud je pozorovaných geodetických cest mezi daným párem uzlů více, připočte se do mezilehlosti uzlu *j* jen podíl těchto cest procházejících přes *j*.

mu mít dvě vazby), nýbrž který je výhodně umístěn na řadě spojnic mezi ostatními, díky čemuž má kontrolu nad toky zdrojů a informací proudících v síti. Zohlednění orientace vazeb výpočet v zásadě nikterak nemění (Borgatti et al. 2013, s. 175). To ovšem neplatí pro vazby ohodnocené. Jedním způsobem, jak obecně řešit analýzu ohodnocených vazeb, je dichotomizace jejich hodnot, tj. transformace na vazby binární. Tím ovšem dochází ke ztrátě informace a zároveň volba prahu pro dichotomizaci do analýzy vnáší jistou arbitrárnost. Jedním etablovaným způsobem, jak lze zohlednit sílu vazeb při výpočtu mezilehlosti, je užití tokové mezilehlosti (*flow betweenness* [Freeman, Borgatti a White 1991]), což je způsob výpočtu založený na možných tocích v síti, přičemž uzel umístěný na vysokém počtu toků má vysokou mezilehlost.

Blízkost

Další populární mírou centrality vycházející z geodetických vzdáleností je blízkost (*closeness*). Výchozím bodem pro blízkost je úvaha, že uzel, který je blízko (tzn. má nízký součet délky geodetických vzdáleností) k ostatním uzlům v síti, je výhodně umístěn pro šíření či získávání informací, zdrojů a vlivu. Součtem geodetických vzdáleností každého uzlu ke všem ostatním vzniká míra, které se také někdy říká vzdálenost (*farness*) a která má tu vlastnost, že nižší hodnoty znamenají centrálnější uzly (neboť čím nižší součet vzdáleností má daný uzel, tím blíže je ostatním). Tato interpretace je však značně neintuitivní a z mnoha ohledů (např. pro statistickou analýzu) nepraktická. Proto se v praxi běžně používá převrácená hodnota součtu geodetických vzdáleností od uzlu *i* k ostatním, která již má žádoucí interpretaci (tj. vyšší hodnota indexu indikuje vyšší centralitu). Formálně pro uzel *i*:

$$c_i = \frac{1}{\sum_j d(j,i)}$$

Blízkost se na první pohled jeví jako konceptuálně i výpočetně jednoduchá metrika. Při bližším prozkoumání má však několik úskalí (Borgatti et al. 2013). Zásadním problémem především z matematického hlediska je, že geodetické vzdálenosti nejsou definovány pro uzly, mezi nimiž není žádná cesta, tj. patří do rozdílných komponent sítě (rozdílných vzájemně nepropojených částí této sítě). To je i situace izolovaných uzlů, které nemají cestu k žádným dalším uzlům v síti, a tak jsou jejich vzdálenosti nedefinované. Toto u mezilehlosti nevadí – není-li možné nijak dosáhnout od jednoho uzlu ke druhému, žádný jiný uzel mezi nimi nehraje roli prostředníka a do jeho skóru mezilehlosti se prostě jen nezapočítají. Jinak je tomu u blízkosti, která k výpočtu vyžaduje sečtení vzdáleností mezi všemi páry uzlů, což je u nedefinovaných vzdáleností pochopitelně problém. Tento problém lze obejít tím, že se vzdálenostem mezi vzájemně nedosažitelnými dvojicemi předem určí fixní hodnota, například nejdelší vzdálenost mezi propojenými uzly zvýšená o jedna. Nebo se mohou izolované uzly z analýzy odstranit. To však jednak komplikuje interpretaci a také zavádí arbitrárnost do celé analýzy. Další problém související s blízkostí je její rozšíření na orientované vazby. Teoreticky se jedná o přímočaré rozlišení na příchozí a odchozí blízkost stejně jako v případě stupně. To

ovšem předpokládá, že lze v empirických sítích rozlišit cesty, na nichž mají všechny vazby stejný směr, což se v praxi takřka nikdy nevyskytuje, neboť cesty jsou zcela běžně složené z vazeb obou směrů.[7] Provizorním řešením je symetrizace vazeb, tedy odstranění jejich orientace. To ovšem podobně jako u dichotomizace ohodnocených vazeb vede ke ztrátě informace a konceptuálním komplikacím v analýze.

Pro všechny tři výše představené míry centrality platí, že svého teoretického maxima nabývají v tzv. grafu či síti hvězdy (*star graph* [Borgatti a Everett 2006; Freeman 1979]). V síti hvězdy je jeden centrální uzel, který má vazby ke všem ostatním uzlům, přičemž ty již mezi sebou navzájem žádné další vazby nemají. V takové síti je tedy centralita maximálně koncentrována, čehož se využívá při výpočtu centralizace a při normalizaci měr centrality. Centralizace je vlastností sítě jako celku[8] a odkazuje na to, do jaké míry je centralita v dané síti koncentrována kolem několika málo uzlů, či zda je naopak rovnoměrně rozložena napříč všemi uzly v síti. Obecně lze hodnotu centralizace podle Freemana (1979) chápat jako míru, do níž se daná síť podobá síti hvězdy. Centralizace nabývá hodnotu 1, pokud je daná síť skutečně sítí hvězdy, zatímco hodnota 0 značí perfektně rovnoměrné rozložení dané míry centrality v síti, a tedy naprostou decentralizaci. Normalizace jednotlivých měr centrality se zpravidla provádí vyjádřením jejich proporce z maximální teoreticky možné hodnoty v dané síti. Například uzel, který má stupeň o hodnotě osm v jedenáctičlenné binární neorientované síti, má normalizovaný stupeň o hodnotě 0,8, protože teoretickým maximem stupně je v tomto případě 10 (tedy vazby ke všem ostatním uzlům). Normalizace je užitečná třeba k porovnávání centrality uzlů ve dvou nestejně velkých sítích, neboť umožňuje porovnat, jak centrální jsou uzly v rámci svých sítí.

Existuje velké množství dalších měr centrality.[9] Trojice výše představených je však zdaleka nejvíce užívána ve výzkumu a je považována za kanonické míry, přičemž mnoho dalších měr centrality je od nich nějakým způsobem odvozených (Borgatti 2005; Borgatti a Everett 2006; Freeman 1979; Landherr et al. 2010; Robins 2015). Zde jsem chtěl poukázat na dvě věci. Jednak na vzájemnou komplementaritu měr centrality – to, že každá z měr centrality operacionalizuje tento koncept mírně jiným způsobem, naznačuje, že nemusí být nadbytečné provést ve výzkumu více výpočtů, neboť výsledné informace se nemusí nutně překrývat a mohou naopak odhalit pozoruhodný aspekt analyzované sítě, což demonstruje graf č. 1. Například ve výzkumu kriminálních sítí jsou pozoruhodní tzv. strategicky umístění aktéři (Morselli 2010) s vysokou mezilehlostí a nízkým stupněm, neboť tito aktéři jsou díky nízkému počtu přímých vazeb méně viditelní, ale zároveň mají díky vysoké mezilehlosti značnou kontrolu nad významnými toky v síti. Dále, přestože jsou výše představené míry často užívané a nejvíce propracované, stále mají několik problémů, jako je jejich nesamozřejmé rozšíření na nebinární, orientovaná či multiplexní nebo bipartitní data (o nichž bude řeč níže). Všechny tři představené míry centrality jsou nicméně plně implementovány

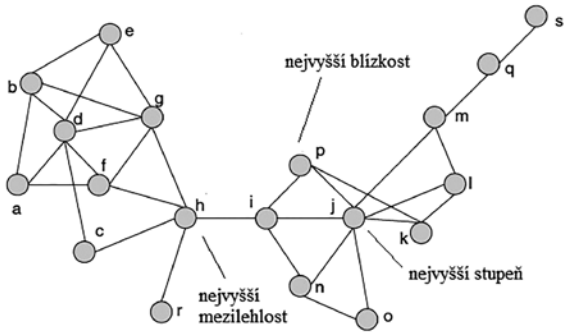
7 V řeči teorie grafů tento předpoklad znamená, že jsou uzly tzv. silně propojené. Prakticky však nalézáme v zásadě jen slabě (tedy vazbami v obou směrech) propojené uzly.

8 Zatímco centralita je vlastností jednotlivých uzlů. Je proto důležité tyto dva pojmy nezaměňovat.

9 Schoch tuto skutečnost výmluvně ilustruje tím, že sestavil „periodickou tabulku“ měr centrality, která je dostupná zde: http://schochastics.net/sna/periodic.html.

v dostupných softwarových balících[10] pro SNA včetně řady jejich úprav a variant, což je spolu s tím, že se váží k řadě zajímavých síťových výzkumných konceptů, činí nejen velmi populárními, ale také často užívanými bez hlubší reflexe či s řadou navazujících chyb. Ty nejzávažnější představíme v následující sekci.

Graf 1: Síť se třemi výše ilustrovanými měrami centrality (převzato podle Ortiz-Arroyo 2010)



Problémy měr centrality

V této části se podíváme na časté problémy, které provázejí využití měr centrality v aplikovaném výzkumu sítí. V zásadě lze tyto problémy seskupit do tří kategorií. Jsou jimi problémy s operacionalizací konceptu centrality a manipulací („hackingem“) měr centrality, problémy související s často pozorovaným kladně šikmým rozdělením měr centrality a konečně problém statistické interdependence a modelování měr centrality. Po představení a reflexi těchto problémů se v další části tohoto textu podíváme na možnosti jejich řešení a prevence.

Problematika operacionalizace

Co se prvního problému, nevalidní operacionalizace a případného „hackování“ měr centrality týče, je nejprve důležité si uvědomit, co centralita představuje. Jak jsme již definovali v předchozí sekci, centralita coby vlastnost uzlů v síti je teoretickým konceptem vyjadřujícím relativní význam daného uzlu v rámci dané sítě. To znamená, že podobně jako jiné teoretické koncepty je třeba jej před samotným měřením operacionalizovat. A jak ukazují míry centrality popsané výše, operacionalizovat centralitu lze různě, neboť sám tento teoretický koncept má různé dimenze – např. dimenzi aktivity, dimenzi přemosťování nebo dimenzi blízkosti (srov. Borgatti 2005; Landherr et al. 2010; Robins 2015; Schoch 2018). Podle toho, která z významových rovin tohoto konceptu nejlépe odpovídá konkrétní výzkumné otázce, by měla

10 Například UCINET (Borgatti, Everett a Freeman 2002), knihovny jazyka R zvané igraph (Csardi a Nepusz 2006) a statnet (Handcock, Hunter, Butts, Goodreau a Morris 2003) či Pajek (Batagelj a Mrvar 1996).

být zvolena i míra centrality. Nestane-li se tak, pak hrozí, že výsledky analýz a jejich interpretace nebudou validní. Kupříkladu zajímá-li nás, kdo je v pracovním kolektivu nejaktivnější, čemuž by nejvíce odpovídal výstupní stupeň, ale místo toho měříme tento koncept prostřednictvím mezilehlosti, která s aktivitou nemusí vůbec nijak souviset, výsledky analýzy jednoduše nemají vzhledem k dané výzkumné otázce žádnou výpovědní hodnotu.

Dalším potenciálním zdrojem podobného pochybení je teoreticky sice validní operacionalizace, která ale nerespektuje konkrétní povahu dat. Můžeme tak například chybně aplikovat míry určené pro nevážené vazby na sítě s vazbami váženými apod. Extrémním případem je pak zvolení (či dokonce vytvoření) specifické míry centrality tak, aby co nejlépe odpovídala snadno interpretovatelným či žádoucím výsledkům. Jak jsme již totiž uvedli, měr centrality je velké množství, a je tak možné, že při opakované analýze dané sítě mnoha různými měrami centrality dojdeme nakonec k takovému závěru, který konvenuje představě o žádoucím výsledku. Tato situace se nápadně podobá často kritizovanému jevu zvanému „p-hacking“ v klasické statistické analýze dat (srov. Soukup 2019). Schoch (2018) například ukázal, že i pro velice jednoduchou a počtem uzlů malou síť florentinských rodů (Padgett a Ansell 1993)[11] lze vytvořit celou řadu měr centrality, které by vygenerovaly takové seřazení rodů podle centrality, jež by bylo kontradiktorní výsledkům původní analýzy. Zde je nutné poznamenat, že nejasnosti ohledně konceptuálních a matematických základů měr centrality takovéto situaci poněkud nahrávají, neboť neexistuje jednoznačný konsenzus ohledně toho, jaké podmínky by měly míry centrality splňovat (Borgatti a Everett 2006). Landherr a kol. (2010) dokládají, že ani výše popsané „kanonické“ míry centrality nemají některé matematické vlastnosti (např. to, že při zkrácení vzdálenosti uzlu *a* k uzlu *b* se centralita *a* nesníží), které by umožnily formulaci univerzálních pravidel pro výběr či konstrukci měr centrality.

Rozdělení centrality a jeho příčiny

Druhým problémem souvisejícím s mírami centrality je otázka tzv. bezškálového (*scale-free*) či mocninného rozdělení centralit a s ním souvisejícím mechanismem preferenční náklonnosti (*preferential attachment* [Barabási a Albert 1999]). Rozdělení měr centrality představují jeden z významných způsobů, jak posoudit strukturu sítě jako celku obzvlášť ve vztahu k míře její centralizace. Za prominencí obou těchto termínů a jejich častým výskytem jak v odborné, tak v populárně-naučné literatuře patrně stojí síťový fyzik Barabási a jeho tým. Především Barabási působí jako velmi aktivní a poutavý popularizátor síťového přístupu a potažmo síťové vědy jako takové (viz např. Barabási 2005). Bezškálové (někdy též mocninné, *power-law*) rozdělení je typem statistického rozdělení spojitých veličin, které se vyznačuje silnou kladnou šikmostí. Toto rozdělení má být podle Barabásiho charakteristické pro míry centrality v sítích, zejména pro stupeň. V praxi to znamená, že by v sítích měly být míry centrality rozděleny tak, že většina přítomných uzlů má jen nízký počet vazeb, zatímco několik málo uzlů má disproporčně vysoký počet vazeb (viz graf č. 2). Řada empirických

11 Tato síť se díky své popularitě, poměrně jednoduché struktuře a snadno pochopitelnému kontextu stala často využívaným příkladem pro demonstraci metodologických principů a nových metod v SNA.

studií napříč specifickými kontexty (sociální, biologické či technologické sítě) pak dochází k závěrům, že pozorované sítě tento typ rozdělení vykazují (Clauset, Shalizi a Newman 2009; Robins 2015). Může se tedy zdát, že bezškálové rozdělení je téměř zákonem sítí ve fyzikálním slova smyslu. Barabási a Albert (1999) k tomu navíc připojili vysvětlení, proč tomu tak je. Tím má být působení mechanismu preferenční náklonnosti, který postuluje, že počet vazeb, jež daný uzel má, působí jako jakýsi magnet pro navázání dalších vazeb, neboť zvyšuje pravděpodobnost, že nově příchozí uzly a nově vytvořené vazby budou směřovat k již centrálním uzlům, které mají být díky svému vysokému počtu více viditelné, snáze dostupné a třeba také výhodnější coby potenciální kontakty. To v posledku vede ke kumulaci vazeb centrálními uzly, jejímž výsledkem má být právě silná šikmost rozdělení jejich počtu.

Graf 2: Porovnání bezškálového a normálního rozdělení počtu vazeb

V tomto ohledu se nejedná o nikterak nový objev, neboť na toto rozdělení a jeho potenciální spojitost s mechanismem akumulace vazeb (nazvaným akumulace výhod) přišel již v 70. letech de Solla Price (1976) v kontextu bibliometrie. V sociologickém kontextu je pak nutné poznamenat, že se v podstatě jedná o síťovou podobu Matoušova efektu (Merton 1968). Častý výskyt kladně sešikmených rozdělení měr centrality spolu s intuitivním mechanistickým vysvětlením pomocí preferenční náklonnosti však nejsou tak mocné, jak by se mohlo zdát, a to ze dvou primárních důvodů. Prvním z nich je, že řada studií vykazujících bezškálové rozdělení ve skutečnosti žádným způsobem netestuje shodu empirického rozdělení s jeho teoretickým předobrazem a mnohdy na základě pouze návodných indikátorů (histogram či index centralizace) usuzují, že se jedná o bezškálové rozdělení. Jak se ale ukazuje, při rigorózním testování ve většině případů nelze prokázat, že by se skutečně jednalo o bezškálové rozdělení a ne jen o rozdělení s vyšší mírou šikmosti (Broido a Clauset 2019; Robins 2015). Co se preferenční náklonnosti týče, tak vyvozovat působení tohoto mechanismu z výsledku, který má generovat, je samo o sobě ukvapeným závěrem, protože ani přítomnost perfektního bezškálového rozdělení ještě neznamená, že nebylo navozeno působením jiného mechanismu nebo spolupůsobením jiných mechanismů. Sítě obecně se vyznačují nelinearitou a komplexitou, kdy se v jedné síti často těžko předvídatelně vzájemně ovlivňuje více souběžně působících mechanismů (Robins, Pattison a Woolcock 2005; Snijders a Steglich 2015). V praxi se pak ukazuje, že při kontrole dalších mechanismů (např. uzávěra či homofilie) může být v řadě případů efekt preferenční náklonnosti neprůkazný či slabý. Lze tedy konstatovat, že ač jsou kladně šikmá rozdělení stupňů (příp. dalších měr centrality)

empirickou pravidelností, nelze je jednoduše označovat za rozdělení bezškálové a vysvětlovat je automaticky pomocí preferenční náklonnosti (Snijders 2013).

Statistické modelování centrality

Snaha hlubším způsobem vysvětlovat, co stojí za měrami centrality (např. pomocí mechanismu preferenční náklonnosti), je spojená s jejich posledním velkým problémem, který si představíme nyní. Míry centrality samy o sobě jsou deskriptivními indikátory, podobně jako jsou míry centrální tendence (průměry či medián) popisnými nástroji ve standardní statistické analýze. To znamená, že míry centrality nejsou modelem a nevysvětlují, jak a proč jsou konkrétní hodnoty či jejich rozdělení takové, jaké jsou v daném případě. Ačkoliv je deskripce zdrojem cenných poznatků, k formulaci hodnověrné explanace nestačí, a není vhodná ani k zohlednění vlivu náhody nebo jiných intervenujících faktorů. Nabízí se tak možnost využít míry centrality jako závisle proměnné v rámci statistického modelu, jako je třeba vícenásobná regresní analýza. To je sice v rovině vědeckého vysvětlování krokem správným směrem, z hlediska statistické inference se však jedná o potenciálně závažnou chybu. Standardní statistická inference totiž staví na zásadním předpokladu, kterým je vzájemná nezávislost pozorování. Jinými slovy, tento předpoklad říká, že hodnoty dané proměnné by měly být nezávislé na hodnotách této proměnné u ostatních pozorování. Tento předpoklad je však v síťové analýze fundamentálně narušen (srov. Robins 2013; Snijders 2011), neboť uzly a vazby jsou v sítích vzájemně interdependentní, což někteří dokonce považují za hlavní důvod, proč sítě vůbec studovat (Brandes, Robins, McCranie a Wasserman 2013).

U měr centrality se interdependence dobře ilustruje. Představme si situaci, kdy jednomu uzlu pozměníme stupeň například tím, že jej snížíme o jednotku odebráním jedné vazby. Toto odebrání vazby ale nepozmění jen stupeň daného uzlu, nýbrž i stupeň uzlu na druhém konci této vazby. S ostatními měrami je pak situace ještě komplikovanější, protože vliv odebrání jedné vazby ovlivňuje jejich hodnoty nejen u uzlů, jež daná vazba spojuje, ale potenciálně i centralitu dalších uzlů v síti. Tato interdependence uzlů hrubě zkresluje směrodatné chyby v jakémkoliv modelování, což má za následek nemožnost korektní statistické inference. Jako alternativu klasické statistické inference lze v takovém případě užít inferenci založenou na randomizaci, která je implementována např. v softwarovém balíku UCINET (více např. v Good 2005). Randomizace však celou situaci řeší pouze částečně. Znáhodněním závisle proměnné se sice odstíní problém interdependence pozorování tím, že se pomocí mnoha set či tisíců permutací vygeneruje distribuce možných výsledků, kterou lze následně použít pro závěry o statistické významnosti. Nicméně to, že se interdependence odstíní, také znamená, že nelze strukturu sítě samotnou zohlednit při modelování, takže dochází ke ztrátě mnohdy zásadní informace.

Možná řešení

V této sekci nastíníme potenciální řešení problémů měr centrality představených v sekci předchozí. Jedná se o typologii toků a k ní vhodným mírám centrality, poziční přístup k analýze sítí a explicitně síťové statistické modely. S výjimkou prvně jmenovaného tato řešení pokrývají více problémů než jen jeden izolovaně.

Typologie toků v síti

Jedním ze způsobů, jak pomoci tomu, aby bylo užití měr centrality co nejvíce validní, je více propojit jejich výpočetní rámec s jejich konceptuálním podložím. Jinak řečeno, je nutné najít vlastnosti sítí a centrálních aktérů, kterým výpočet dané míry centrality nejlépe odpovídá. Borgatti (2005) v této souvislosti navrhl typologii vazeb podle toho, jaký proces či tok reprezentují (např. směnu, komunikaci, interakci či dlouhodobé vztahy), a podle cest, jaké mohou tyto procesy v síti sledovat, tj. zda je smysluplné hovořit o tom, zda by se např. aktéři v síti mohli snažit vzdálenosti mezi sebou optimalizovat. Borgatti dochází k tomu, že toky v sítích lze rozlišit na tři typy: paralelní duplikaci, seriální duplikaci a transfer. Paralelní duplikace znamená, že to, co po vazbách proudí, se kopíruje od jednoho uzlu ke druhému a každá taková duplikace může probíhat současně s dalšími. Příkladem je zde elektronická komunikace prostřednictvím chatu nebo e-mailu, kdy jednu zprávu lze posílat a následně přeposílat libovolnému množství příjemců naráz. Seriální duplikace je také založená na tom, že se daný tok kopíruje, nicméně nešíří se paralelně, ale postupně vždy jeden uzel za druhým, a nikoliv několik najednou. Typickým příkladem je zde šíření drbů či poskytování podpory, ať už instrumentální, nebo afektivní. Takové toky mohou proudit po síti libovolným způsobem a skončit nakonec u mnoha uzlů zároveň, leč vždy se šíří jen v rámci jedné vazby. Posledním typem toků je transfer, kdy po síti proudí konkrétní materiální objekt (např. peníze či zboží). Uskutečněním transferu jeho odesilatel ztrácí nad daným objektem kontrolu, kteroužto přebírá příjemce. Transfery tedy postupují vazbu za vazbou, na rozdíl od seriální duplikace však jejich obsah není kopírován (jako jsou drby), nýbrž předáván.[12]

Druhým kritériem toků v síti, jsou cesty, které využívají. V teorii grafů lze rozlišit tři typy cest: sledy (*walk*), tahy (*trail*) a cesty (*path*). Sledy jsou takové sekvence uzlů a vazeb, v nichž se mohou uzly i vazby libovolně opakovat. Tahy jsou takové sledy, v nichž se mohou opakovat uzly, ale neopakují se v nich žádné vazby. A konečně v cestách se neopakují ani uzly, ani vazby. Zvláštním případem jsou pak již výše zmíněné geodetické cesty, tj. nejkratší cesty mezi danou dvojicí uzlů. Tradičně se při měření centrality uvažuje v intencích geodetických cest, neboť to umožňuje jednoznačné a výpočetně únosné kalkulace, což byl pádný důvod především před masovým rozšířením výkonných počítačů. A u takových toků, kde se jedná o transfery s předem daným cílem, je možné jako vyjádření optimalizace využít geodetické vzdálenosti. Jindy je však žádoucí zohlednit i jiný typ spojení mezi uzly. Například bankovka se po síti může pohybovat zcela libovolně a ocitnout se i několikrát v rukou jednoho aktéra, a po síti tak cestuje prostřednictvím sledů. Oproti tomu pikantní drb cestuje pravděpodobně po tazích, neboť nemusí nutně sledovat nejkratší spojnice a může se klidně dostat k jednomu aktérovi z více různých zdrojů, leč pravděpodobně nemá smysl

12 Borgattiho (2005) typologie se soustředí na problematiku aktérů a vazeb mezi nimi, což je situace typická pro většinu společenskovědních aplikací SNA. V tomto ohledu zůstává nezodpovězena otázka, jak pojmout vazby, které nemají povahu toků. To je relevantní například v lingvistice či psychologii, kde uzly mohou reprezentovat entity, které nemají agentickou povahu (tedy nejsou to aktéři, např. slova, koncepty či konstrukty), a vazby mohou reprezentovat jejich blízkosti či podobnosti. V tomto kontextu totiž nemusí nutně mít smysl uvažovat o tocích prostřednictvím vazeb.

uvažovat o jeho opakování (od *a* k *b*) či reciprokém sdělení téhož drbu (od *b* k *a*). Celou tuto situaci s dalšími konkrétními příklady zachycuje tabulka č. 1. Smyslem celé této Borgattiho úvahy je možnost systematicky se zamyslet, jakým způsobem převádíme podstatné vlastnosti zkoumaných jevů do řeči teorie grafů, a vzhledem k tomu vybrat vhodnou míru centrality tyto vlastnosti adekvátně zachycující.

Tabulka 1: Typologie toků v síti (Borgatti 2005)

	paralelní duplikace	**seriální duplikace**	**transfer**
geodetické cesty	žádný proces	mitotická reprodukce	doručení zásilky
cesty	jmenné servery online	virová infekce	žebrání
tahy	posílání e-mailů	drb	použité zboží
sledy	ovlivňování postojů	emoční podpora	směna peněz

Poziční přístup

Borgattiho typologii toků a cest lze vnímat jako jeden ze způsobů, jak v rámci stávajícího výzkumu konceptuálně napravit nedostatky spojené s užíváním měr centrality. Komplexnější způsob, jak k tomuto problému přistoupit, představuje poziční přístup (*positional approach*) k analýze sítí navržený Brandesem (2016). Poziční přístup lze v jistém slova smyslu vnímat jako „neparametrický“ přístup k analýze sítí, neboť má za ambici redefinovat klasické síťové metody na jednotícím matematickém základu bez jakýchkoliv zamlčených předpokladů. Poziční přístup by měl do budoucna redefinovat všechny známé síťové koncepty, doposud je ovšem omezen (nejspíše pro svou novost) na míry centrality, kde však již tato perspektiva přinesla nové poznatky a možnosti. Jak již název napovídá, jádrem pozičního přístupu je pojem síťové pozice (*network position*), který má být výchozím bodem pro definici všech ostatních konceptů. Síťová pozice je vlastně umístěním uzlu v multidimenzionálním prostoru definovaném vlastnostmi uzlů (atributy) a jejich vazbami (Brandes 2016). Toto umístění přiřazuje každému uzlu v síti určité souřadnice na každé z daných dimenzí. Brandes síťovou pozici záměrně definuje takto otevřeně, aby umožňovala zahrnutí různých typů vazeb i různých typů atributů a byla tak co nejflexibilnější. Pro sociální vědy specificky by síťová pozice měla být extenzí pozice v sociálním prostoru, tedy například jeho pozice v rámci sociální stratifikace (Blau 1977; McPherson 1983 podle Brandes 2016). Síťové pozice by totiž, podobně jako individuální pozice, mělo být možné vzájemně porovnávat, a tedy říci, který uzel je postavený výše a který níže v konkrétní dimenzi. Z této myšlenky pak lze odvodit tzv. poziční dominanci (*positional dominance*), tedy situaci, kdy uzel *i* je alespoň na jedné dimenzi umístěn výše než uzel *j* a zároveň na žádné jiné není umístěn níže. Do této definice lze zahrnout míry centrality, ale i další charakteristiky, které v posledku umožní seřadit uzly v síti podle poziční dominance.

Schoch a Brandes (2016) v rámci pozičního přístupu ukazují, že jednotící podmínkou měr centrality je předuspořádání inkluze sousedství (*neighbourhood inclusion preorder*). Toto předuspořádání říká, že má-li uzel *a* vazbu na všechny uzly jako uzel *b* a zároveň má více vazeb celkově, pak uzel *a* nemůže být nikdy méně centrální než uzel *b*. Pokud má

například uzel *a* vazbu na uzly *b*, *c* a *d*, zatímco uzel *b* jen na *c* a *d*, pak *a* nemůže být méně centrální než *b*, neboť jeho sousedství zahrnuje sousedství uzlu *b*. Tato vlastnost je klíčová, neboť jen ty míry centrality, které takovou podmínku splňují (což jsou mj. všechny tři představené v úvodu), jsou reliabilní a odolné vůči manipulaci. Schoch a Brandes (2016) proto doporučují konstruovat nové míry centrality tak, aby zachovávaly předuspořádání inkluze sousedství. Vzhledem k tomu, že se poziční přístup explicitně snaží o rigorózní matematické redefinování síťových konceptů, jedná se o matematicky velice komplexní přístup, jehož celková expozice značně přesahuje rámec této stati. Nicméně konkrétní aplikace pozičního přístupu pro identifikaci centrálních aktérů lze představit i bez takto komplexního výkladu. Těmito aplikacemi je analyzování centrality uzlů bez konkrétních hodnot měr centrality prostřednictvím určování jejich pořadí v rámci sítě nebo určování pravděpodobnosti, s jakou bude daný uzel mít konkrétní příčku v rámci takového seřazení (Schoch 2018). Schoch (2018) vychází z analýzy několika příbuzných měr založených na vzdálenostech mezi uzly coby výchozí souřadnice definující pozice jednotlivých uzlů. Prostřednictvím těchto souřadnic pak definuje relace poziční dominance pro každou dvojici uzlů, přičemž ne vždy je jednoznačně možné stanovit, který uzel dominuje nad kterým (např. jsou-li v naprosto ekvivalentních pozicích nebo nemají-li žádný průnik sousedství a zároveň stejné hodnoty jednotlivých výchozích souřadnic). Neporovnatelnost či stejná pořadí pak vnášejí do pořadí uzlů určitou míru nejistoty, kterou lze vyjádřit pravděpodobnostně, což de facto umožňuje říci, jak je pravděpodobné, že jeden uzel bude řazen výše než jiný. Díky tomu tak lze pozičním přístupem zachytit v měření centrality i aspekt náhody tak, jak jsem naznačil výše u problémů s interdependencí a nejistotou.

Statistické modely pro síťová data

S interdependencí spjatou se síťovými daty a nejistotou souvisí statistické modelování sítí. Připomeňme, že statistické modelování sítí je zásadně odlišné od klasického statistického modelování a že prosté odstranění vlivu interdependence randomizací není postačující. V posledních zhruba deseti až patnácti letech proto dochází na poli síťových modelů k prudkému rozvoji (Snijders 2011). Patrně nejznámější a nejpoužívanější z modelů explicitně navržených pro síťová data jsou tzv. exponenciální modely náhodných grafů (*exponential random graph model,* též ERGM [Lusher, Koskinen a Robins 2013]). Tyto modely či modely jim podobné modelují strukturu sítě prostřednictvím tzv. konfigurací, což jsou strukturní elementy sítě, např. trojúhelníky či reciproké vazby, jejichž vliv na tvorbu vazeb v modelované síti je posouzen prostřednictvím počítačových simulací. ERGM vycházejí z předpokladu, že konfigurace reprezentují mechanismy, jejichž působením vzniká struktura sítě. Například v na první pohled silně centralizované síti může analýza pomocí ERGM dojít k závěru, že centralizace nebyla způsobena efektem preferenční náklonnosti (reprezentovanou konfigurací hvězdy), nýbrž efektem triadické uzávěry (reprezentované konfigurací trojúhelníku), a tedy že se jedná o vedlejší účinek tendence aktérů uzavírat trojúhelníky spíše než o jejich tendenci akumulovat vazby. Díky tomu, že ERGM modelují strukturu sítě z konfigurací, dokážou adekvátně posoudit povahu interdependence mezi jednotlivými vazbami a zohlednit přitom také náhodné kolísání.

ERGM však modelují strukturu, zatímco modelování centrality aktérů by vyžadovalo modelování individuálních atributů, resp. jejich distribuce nebo pořadí. V tomto ohledu se negativně projevuje relativní „mládí“ síťových modelů, neboť v současnosti nám takový model, a především jeho praktická implementace chybí. Krok v tomto směru představuje tzv. autologistický model atributů aktérů (*autologistic actor attribute model*, též ALAAM [Robins, Pattison a Elliott 2001]). Tento model je výpočetně založen na stejných simulacích jako ERGM, jen jako závisle proměnnou využívá individuální atribut a vzhledem k tomu vychází z odlišných konfigurací. Problém je ale v tom, že ALAAM je zatím navržen a implementován jen pro vysvětlování dichotomických závisle proměnných. ALAAM lze přirovnat k binární logistické regresi, která navíc zohledňuje interdependenci mezi uzly a vazbami v síti. Podobně jako u binární logistické regrese je konceptuálně možné rozšířit ALAAM na modelování ordinálních či diskrétních atributů, jako je tomu třeba u ordinální logistické regrese nebo negativní binomické regrese, leč k těmto rozšířením zatím nedošlo ani konceptuálně, natož v rámci implementace do přístupného analytického softwaru. Prudký rozvoj modelů spolu s obnovenou diskusí o měření centrality v posledních letech však dává naději, že se podobného modelu v blízké budoucnosti dočkáme.

Bipartitní a multiplexní sítě

Doposud jsme se v této studii omezili pouze na nejběžnější a výchozí případ sítí s jedním typem uzlů a jedním typem vazeb mezi nimi, který lze také označit jako unipartitní a uniplexní. V této části se stručně podíváme na specifika měr centrality, která souvisejí s jejich aplikací na méně obvyklé typy dat, jakými jsou bipartitní a multiplexní sítě.

Bipartitní či bimodální (*two-mode*) sítě zachycují dva odlišné nepřekrývající se typy uzlů (módy), přičemž vazby připouštějí pouze napříč módy, a nikoliv uvnitř nich, tj. mezi uzly stejného módu. Typicky jeden mód reprezentuje aktéry, zatímco druhý jejich různé afiliace či skupiny a vazby pak představují členství aktérů v afiliacích/skupinách. Příkladem mohou být politici a jejich členství v různých skupinách a výborech, studenti docházející na různé semináře nebo interpreti vystupující na různých jevištích. Pro bipartitní síť je samozřejmě nutné většinu konceptů redefinovat tak, aby dávaly smysl. To platí i pro centralitu a její měření. Zatímco dřívější praxe byla převést bipartitní síť pomocí různých technik projekce na dvě běžné (unimodální či unipartitní) sítě reprezentující síť společných afiliací mezi aktéry a síť překryvu aktérů mezi afiliacemi (srov. Borgatti a Everett 1997), v současnosti se má tento způsob analýzy za překonaný, neboť vede ke ztrátě informace o některých aspektech původní sítě. Proto byl vyvinut přístup tzv. duální projekce (*dual-projection* [Everett a Borgatti 2013]), který umožňuje analyzovat bipartitní sítě sice složitějším, ale ke ztrátě informací nenáchylným způsobem, pročež jedním z jeho využití je právě i identifikace centrálních aktérů, která umožňuje korektně posoudit centralitu uzlů v obou módech najednou, a tedy i v dané síti celkově (Everett 2016).

Další z výzev pro měření centrality představují tzv. multiplexní sítě, což jsou sítě, které zachycují několik odlišných typů vazeb mezi jednou množinou uzlů (Kivelä et al. 2014). Příkladem zde je analýza a srovnávání afektivních a instrumentálních vazeb, jakými jsou například přátelství a pomoc s úkoly mezi žáky ve třídě nebo svěřování se a spolupráce mezi

zaměstnanci jedné firmy. Ani u multiplexních sítí nelze jednoduše převzít míry centrality z běžných (v tomto případě uniplexních či simplexních) sítí a aplikovat je bezprostředně na sítě multiplexní. Podobně jako u bipartitních sítí, dřívější praxí bylo pomocí různých metod (např. sjednocení či součet) agregovat všechny typy vazeb do jednoho všeobjímajícího typu, což je ale provázeno nejen ztrátou informace, nýbrž i výrazně komplikovanou interpretací výsledků (Hanneman a Riddle 2005). Jedním ze způsobů, jak zachovat původní informaci a přitom zohlednit i aspekt multiplexity, je nejen výpočet měr centrality, ale i výpočet jejich variance napříč různými typy vazeb tak, aby bylo možné odlišit uzly dominantně propojené v jednom typu vazeb od těch, které mají vazby rovnoměrně rozprostřeny napříč všemi zkoumanými typy vazeb (Battiston, Nicosia a Latora 2014). V jiných případech, kdy je analyzovaných typů vazeb větší množství, může být užitečné charakterizovat adekvátní míru/míry centrality ve všech dimenzích a následně aplikovat analýzu hlavních komponent nebo jinou z technik redukce dimenzionality k nalezení hlavních zdrojů variability za centralitou jednotlivých uzlů. Vývoj indexů i dalších možností, jak vyjádřit centralitu v multiplexních sítích, je však stále na svém počátku (Kivelä et al. 2014) a je otevřenou otázkou, co dalšího z něj vzejde.

Závěr

Měření centrality je jedním ze základních úkolů v analýze sítí. Za tímto účelem byla zkonstruována řada měr centrality. V této studii jsme představili tři z nejstarších a nejzákladnějších měr centrality – stupeň, mezilehlost a blízkost. Ačkoliv jsou tyto i další míry centrality snadno uchopitelné a dostupné v různých softwarových balících, provází jejich aplikaci a interpretaci řada situací, v nichž se můžeme při výzkumu dopustit chyb a nepřesností. Freeman (1979) si ve svém vlivném článku posteskl, že o tom, co je centralita, jak ji teoreticky chápat a jak ji měřit, panuje řada nejasností. Zdá se, že Freemanův postřeh platí i po čtyřiceti letech. V této studii jsem poukázal na několik nejpalčivějších problémů, jako je špatná operacionalizace, statistická inference v sítích a často se vyskytující sešikmené rozdělení centrality a jeho příčiny. Také jsem nastínil možná řešení těchto problémů v podobě typologií toků v sítích, pozičního přístupu a nových statistických modelů pro síťová data. Byť zdaleka ne všechna tato potenciální řešení jsou úplná či jednoduchá, ukazují, že existuje prostor pro inovace, které mohou přispět k tomu, že zanedlouho již nemusí Freemanovo tvrzení platit.

Je zřejmé, že v oblasti měření centrality v SNA existuje široká paleta problémů, které ještě nemají jednoznačné řešení. Budoucí výzkum se tak může zaměřit různými směry od) statistického modelování přes teoreticko-metodologické základy měr centrality až po jejich rozšíření na nejrůznější typy dat. V tomto ohledu se otevírá prostor i výzkumníkům a výzkumnicím z českého prostředí. Ti se mohou díky tomu zapojit do metodologického rozvoje a přispět vědeckému poznání ve světovém měřítku.

Literatura

BARABÁSI, Albert-László. 2005. *V pavučině sítí*. Praha: Paseka.

BARABÁSI, Albert-László a Réka ALBERT. 1999. „Emergence of Scaling in Random Networks." *Science* 286(5439): 509–512 (https://doi.org/10.1126/science.286.5439.509).

BATAGELJ, Vladimir a Andrej MRVAR. 1996. *Pajek—Program for Large Network Analysis* (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

BATTISTON, Federico, Vincenzo NICOSIA a Vito LATORA. 2014. „Structural Measures for Multiplex Networks." *Physical Review E* 89: 032804 (https://doi.org/10.1103/PhysRevE.89.032804).

BENJAMIN, Daniel. J., James O. BERGER., Magnus JOHANESSON a kol. 2017. „Redefine Statistical Significance." *Nature Human Behaviour* 2: 6–10 (https://doi.org/10.1038/s41562-017-0189-z).

BLAU, Peter M. 1977. „A Macrosociological Theory of Social Structure." *American Journal of Sociology* 83(1): 26–54.

BORGATTI, Stephen P. 2005. „Centrality and Network Flow." *Social Networks* 27(1): 55–71.

BORGATTI, Stephen P. a Martin G. EVERETT. 1997. „Network Analysis of 2-Mode Data." *Social Networks* 19(3): 243–269.

BORGATTI, Stephen P. a Martin G. EVERETT. 2006. „A Graph-Theoretic Perspective on Centrality." *Social Networks* 28(4): 466–484.

BORGATTI, Stephen P., Martin G. EVERETT a Linton C. FREEMAN. 2002. *UCINET 6 for Windows: Software for Social Network Analysis*. Analytic Technologies.

BORGATTI, Stephen P., Martin G. EVERETT a Jeffrey. C. JOHNSON. 2013. *Analyzing Social Networks*. SAGE publications.

BRANDES, Ulrik. 2016. „Network Positions." *Methodological Innovations* 9: 2059799116630650 (https://doi.org/10.1177/2059799116630650).

BRANDES, Ulrik, Garry L. ROBINS, Anne McCRANIE a Stanley WASSERMAN. 2013. „What is Network Science?" *Network Science* 1(1): 1–15 (https://doi.org/10.1017/nws.2013.2).

BROIDO, Anna D. a Aaron CLAUSET. 2019. „Scale-free Networks are Rare." *Nature Communications* 10: 1017 (https://doi.org/10.1038/s41467-019-08746-5).

CAMERER, Colin F., Anna DREBER, Felix HOLZMEISTER a kol. 2018. „Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour* 2(9): 637–644 (https://doi.org/10.1038/s41562-018-0399-z).

CASTELLS, Manuel. 2009. *The Rise of the Network Society,* 2nd edition. Wiley-Blackwell.

CÍSAŘ, Ondřej a Jiří NAVRÁTIL. 2015. „Promoting Competition or Cooperation? The Impact of EU Funding on Czech Advocacy Organizations." *Democratization* 22(3): 536–559 (https://doi.org/10.1080/13510347.2013.869742).

CLAUSET, Aaron, Cosma R. SHALIZI a Mark E. J. NEWMAN. 2009. „Power-Law Distributions in Empirical Data." *SIAM Review* 51(4): 661–703 (https://doi.org/10.1137/070710111).

CSARDI, Gábor a Tamasz NEPUSZ. 2006. *The Igraph Software Package for Complex Network Research*. Cit. 30 dubna 2021 (http://igraph.sf.net).

DAŇKOVÁ, Hana, Josef BERNARD a Petr VAŠÁT. 2019. „Using the Respondent-Driven Sampling Method to Survey Homeless Populations: Basic Principles, Application and Practical Recommendations." *Czech Sociological Review* 55(2): 189–214 (https://doi.org/10.13060/00380288.2019.55.2.458).

DE SOLLA PRICE, Derek. 1976. „A General Theory of Bibliometric and Other Cumulative Advantage Processes." *Journal of the American Society for Information Science* 27(5): 292–306 (https://doi.org/10.1002/asi.4630270505).

DIVIÁK, Tomáš 2017. „Equivalence and Blockmodeling in the Analysis of Social Networks." *Naše společnost* 15(1): 27–40 (https://doi.org/10.13060/1214438X.2017.1.15.366).

DIVIÁK, Tomáš, Jan K. DIJKSTRA a Tom A. B. SNIJDERS. 2018. „Structure, Multiplexity, and Centrality in a Corruption Network: The Czech Rath Affair.“ *Trends in Organized Crime* 22: 274–297 (https://doi.org/10.1007/s12117-018-9334-y).

DIVIÁK, Tomáš, Jan K. DIJKSTRA a Tom A. B. SNIJDERS. 2019. „Poisonous Connections: A Case Study on a Czech Counterfeit Alcohol Distribution Network.“ *Global Crime* 21(1): 51–73 (https://doi.org/10.1080/17440572.2019.1645653).

EVERETT, Martin. G. 2016. „Centrality and the Dual-Projection Approach for Two-Mode Social Network Data.“ *Methodological Innovations* 9: 2059799116630662 (https://doi.org/10.1177/2059799116630662).

EVERETT, Martin G. a Stephen. P. BORGATTI. 2013. „The Dual-Projection Approach for Two-Mode Networks.“ *Social Networks* 35(2): 204–210 (https://doi.org/10.1016/j.socnet.2012.05.004).

FREEMAN, Linton C. 1979. „Centrality in Social Networks Conceptual Clarification.“ *Social Networks* 1(3): 215–239 (https://doi.org/10.1016/0378-87337890021-7).

FREEMAN, Linton C., Stephen P. BORGATTIA a D. R. WHITE. 1991. „Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow.“ *Social Networks* 13(2): 141–154 (https://doi.org/10.1016/0378-87339190017-N).

GOOD, Phillip I. 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses,* 3rd ed. Springer.

HANDCOCK, Mark S., David R. HUNTER, Carter T. BUTTS, Steven M. GOODREAU a Martina MORRIS. 2003. *Statnet: Software Tools for the Statistical Modeling of Network Data* (http://statnetproject.org).

HANNEMAN, Robert a Mark RIDDLE. 2005. *Introduction to Social Network Methods*. Cit. 5. března 2021 (http://faculty.ucr.edu/~hanneman/nettext/).

KIVELÄ, Mikko, Alex ARENA, Marc BARTHELEMY, James P. GLEESON, Yamir MORENO a Mason A. PORTER. 2014. „Multilayer Networks.“ *Journal of Complex Networks* 2(3): 203–271 (https://doi.org/10.1093/comnet/cnu016).

KLOBUŠIAKOVÁ, Patrícia, Radek MAREČEK, Jan FOUSEK, Eva VÝTVAROVÁ a Irena REKTOROVÁ. 2019. „Connectivity Between Brain Networks Dynamically Reflects Cognitive Status of Parkinson’s Disease: A Longitudinal Study.“ *Journal of Alzheimer’s Disease* 67(3): 971–984 (https://doi.org/10.3233/JAD-180834).

LANDHERR, Andrea, Bettina FRIEDL a Julia HEIDEMANN. 2010. „A Critical Review of Centrality Measures in Social Networks.“ *Business and Information Systems Engineering*, 2(6): 371–385 (https://doi.org/10.1007/s12599-010-0127-3).

LATOUR, Bruno. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory,* 1st edition. Oxford University Press.

LUKE, Douglas. A. a J. K. HARRIS. 2007. „Network Analysis in Public Health: History, Methods, and Applications.“ *Annual Review of Public Health* 28(1): 69–93 (https://doi.org/10.1146/annurev.publhealth.28.021406.144132).

LUSHER Dean, Johan KOSKINEN a Garry L. ROBINS (eds.). 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

MAZÁK, Jaromír a Tomáš DIVIÁK. 2018. „Transactional Activism without Transactions: Network Perspective on Anti-Corruption Activism in the Czech Republic.“ *Social Movement Studies* 17(2): 203–218 (https://doi.org/10.1080/14742837.2017.1404904).

McPHERSON, Miller 1983. „An Ecology of Affiliation.“ *American Sociological Review* 48(4): 519–532 (https://doi.org/10.2307/2117719).

MERTON, Robert K. 1968. „The Matthew Effect in Science: The Reward and Communication Systems of Science are Considered.“ *Science* 159(3810): 56–63 (https://doi.org/10.1126/science.159.3810.56).

MORSELLI, Carlo. 2009. *Inside Criminal Networks*. New York: Springer.

MORSELLI, Carlo. 2010. „Assessing Vulnerable and Strategic Positions in a Criminal Network." *Journal of Contemporary Criminal Justice* 26(4): 382–392 (https://doi.org/10.1177/1043986210377105).

NUZZO, Regina. 2014. „Scientific Method: Statistical Errors." *Nature* 506(7487): 150–152 (https://doi.org/10.1038/506150a).

OCELÍK, Petr, Kamila SVOBODOVÁ, Markéta HENDRYCHOVÁ, Lukáš LEHOTSKÝ, Jo-Anne EVERINGHAM, Saleem ALI, Jaroslaw BADERA a Alex LECHNER. 2019. „A Contested Transition toward a Coal-Free Future: Advocacy Coalitions and Coal Policy in the Czech Republic." *Energy Research a Social Science* 58: 101283 (https://doi.org/10.1016/j.erss.2019.101283).

OPEN SCIENCE COLLABORATION. 2015. Estimating the Reproducibility of Psychological Science." *Science* 349(6251): aac4716 (https://doi.org/10.1126/science.aac4716).

ORTIZ-ARROYO, Daniel. 2010. „Discovering Sets of Key Players in Social Networks." Pp. 27–47 in Ajith ABRAHAM, Aboul-Ella HASSANIEN a Václav SNÁŠEL (eds.). *Computational Social Network Analysis*. London: Springer.

PADGETT, John F. a Christopher K. ANSELL. 1993. „Robust Action and the Rise of the Medici, 1400–1434." *American Journal of Sociology* 98(6): 1259–1319.

PRELL, Christina. 2011. *Social Network Analysis*, 1st edition. SAGE Publications Ltd.

ROBINS, Garry L. 2013. „A Tutorial on Methods for the Modeling and Analysis of Social Network Data." *Journal of Mathematical Psychology* 57(6): 261–274 (https://doi.org/10.1016/j.jmp.2013.02.001).

ROBINS, Garry L. 2015. *Doing Social Network Research*. SAGE Publications.

ROBINS, Garry L., Philippa PATTISON a Peter ELLIOTT. 2001. „Network Models for Social Influence Processes." *Psychometrika* 66(2): 161–189 (https://doi.org/10.1007/BF02294834).

ROBINS, Garry L., Philippa PATTISON a Jodie WOOLCOCK. 2005. „Small and Other Worlds: Global Network Structures from Local Processes." *American Journal of Sociology* 110(4):894–936.

SCHOCH, David a Ulrik BRANDES. 2016. „Re-Conceptualizing Centrality in Social Networks." *European Journal of Applied Mathematics* 27(6): 971–985 (https://doi.org/10.1017/S0956792516000401).

SCHOCH, David. 2018. „Centrality without Indices: Partial Rankings and Rank Probabilities in Networks." *Social Networks* 54: 50–60 (https://doi.org/10.1016/j.socnet.2017.12.003).

SNIJDERS, Tom A. B. 2013. „Network Dynamics." Pp. 252–280 in Rafael WITTEK, Tom A. B. SNIJDERS a Victor NEE (eds). *The Handbook of Rational Choice Social Research*. Stanford Social Sciences, an imprint of Stanford University Press.

SNIJDERS, Tom A. B. a Christian E. G. STEGLICH. 2015. „Representing Micro–Macro Linkages by Actor-based Dynamic Network Models." *Sociological Methods a Research*, 44(2): 222–271 (https://doi.org/10.1177/0049124113494573).

SNIJDERS, Tom A. B. 2011. „Statistical Models for Social Networks." *Annual Review of Sociology* 37(1): 131–153 (https://doi.org/10.1146/annurev.soc.012809.102709).

SOUKUP, Petr. 2019. „The Use of Statistical and Substantive Significance in the Czech Social Sciences." *Czech Sociological Review* 55(2): 215–254 (https://doi.org/10.13060/00380288.2019.55.2.459).

VALENTE, Thomas W. a Patchareeya PUMPUANG. 2007. „Identifying Opinion Leaders to Promote Behavior Change." *Health Education a Behavior* 34(6): 881–896 (https://doi.org/10.1177/1090198106297855).

VAŠÁT, Petr a Josef BERNARD. 2015. „Formování komunit, nebo sociální integrace? Analýza personálních sítí ukrajinských imigrantů v Plzni." *Czech Sociological Review*, 51(2): 199-226 (https://doi.org/10.13060/00380288.2015.51.2.159).

WASSERMAN, Stanley a Katherine FAUST. 1994. *Social Network Analysis: Methods and Applications*, 1st edition. Cambridge: Cambridge University Press.

Autor

Tomáš Diviák (*1991) působí jako Presidential Fellow na katedře kriminologie a v Mitchell Centre for Social Network Analysis při univerzitě v Manchesteru. Spolupracuje také s katedrou sociologie Filozofické fakulty Univerzity Karlovy a Centrem pro modelování biologických a sociálních procesů. Zabývá se především aplikací analýzy sociálních sítí ve studiu organizovaného zločinu, ale o analýzu sítí se zajímá i metodologicky nebo v souvislosti s její aplikací ve výzkumu veřejného zdraví, politických či historických sítí.
Kontakt: tomas.diviak@ff.cuni.cz