ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

doi: https://doi.org/10.15407/dopovidi2019.06.012
УДК 519.7

О.Д. Поліщук

Інститут прикладних проблем механіки і математики ім. Я.С. Підстригача НАН України, Львів
E-mail: od_polishchuk@ukr.net

Редукція складності моделей мережевих структур та систем

Представлено академіком НАН України Р.М. Кушніром

Аналізується проблема складності мережевих структур та систем. Визначаються кількісні показники складності мережі та їх застосування для вибору ефективної моделі структури системи. Пропонуються методи редукції складності моделей мережевих систем з одночасним відстеженням збереження міри їх адекватності. Ефективність запропонованих підходів ілюструється на прикладах реальних складних систем.

Ключові слова: *мережева система, складність, редукція, серцевина, інкапсуляція, адекватність.*

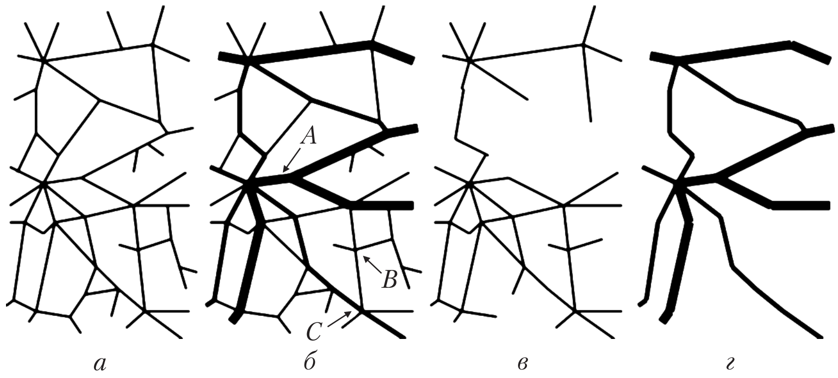
Системи різного походження, типу та призначення є чи не найскладнішим об'єктом наукових досліджень. Це пояснюється необхідністю визначення їх складу та структури, пізнання законів функціонування та особливостей взаємодії між собою та з оточуючим середовищем великої кількості різнорідних об'єктів, які діють для досягнення спільної і часто не до кінця зрозумілої досліднику цілі. Вивчаючи реальні складні системи, ми насправді досліджуємо побудовані на основі спостережень, експериментальних та теоретичних досліджень моделі цих систем (інформаційні, структурні, функціональні, математичні тощо). Одним із напрямків системних досліджень, який почав бурхливо розвиватися протягом останніх десятиліть, стало вивчення складних мережевих систем (СМС), які є чи не найбільш розповсюдженими у мікро- та макросвіті, біологічних системах, людському соціумі [1, 2] тощо. Серед основних особливостей СМС можна назвати наявність у структурі системи великої кількості елементів та їх приналежність до категорії так званих конденсованих середовищ, які можна досліджувати лише в цілому [3]. Визначальною ознакою реально функціонуючих СМС є рух потоків у них. В одних випадках забезпечення руху потоків є основною ціллю утворення та функціонування таких систем (транспортні, фінансові, торговельні та соціальні мережі і т. ін.), у інших — процесом, який забезпечує їх життєдіяльність (рух крові, лімфи, нейроімпульсів у людському тілі тощо). Зупинка руху потоків призводить до припинення існування таких систем.

Показники складності моделей мережевих структур. Структура мережі, тобто сукупність вузлів та поєднуючих їх зв'язків, повністю описується її матрицею суміжності $A=\{a_{ij}\}_{i,\,j=1}^{N}$, у якій для найпростішого випадку бінарних мереж значення $a_{ij}=1$, якщо зв'язок між вузлами n_i та n_j існує, та $a_{ij}=0$, $i,\,j=\overline{1,N}$, якщо такий зв'язок відсутній. Складність моделі мережевої структури насамперед визначається розмірністю N цієї матриці. Іншим, не менш важливим показником складності можна вважати кількість зв'язків у СМС. Побудова моделі мережі розмірності 10^6—10^{12}, тобто заповнення відповідної матриці суміжності, загалом потребує визначення 10^{12}—10^{24} її елементів, що є надскладною проблемою навіть для сучасних суперкомп'ютерів [4]. Очевидно, що дві моделі однакової розмірності матимуть різну коннекційну (від англ. *connection* — зв'язок) складність, якщо кількість зв'язків між їх вузлами є суттєво різною. Більш того, за кількістю зв'язків модель більшої розмірності може бути простішою, ніж модель меншої розмірності. Це пояснюється тим, що чим більше зв'язків у структурі, тим більше варіантів взаємодії елементів системи, а отже тим складнішим може бути процес її функціонування. Очевидно, що збільшення коннекційної складності структури призводить до підвищення перколяційного порогу, зменшення середньої довжини найкоротшого шляху, збільшення у середньому ступеня кожного вузла мережі та коефіцієнта її кластеризації, тобто наближення топології структури до так званих мереж "тісного світу". Це може породжувати як позитивні (збільшення кількості найкоротших шляхів сприяє більш ефективній організації руху потоків мережею, зменшенню середньої завантаженості кожного ребра, розширенню множини альтернативних шляхів руху потоків і т. ін.), так і негативні (пришвидшення поширення інфекційних хвороб та комп'ютерних вірусів тощо) чинники, які впливають на процес функціонування системи. Коннекційну складність моделі визначатимемо за допомогою параметра $\varpi=L/N$, де L — кількість зв'язків мережі (ненульових елементів матриці **A**). Оскільки для орієнтованих мереж об'єми потоків між суміжними вузлами у різних напрямках є різними, то за відсутності петель значення $L\in[2(N-1),N(N-1)]$ та $\varpi\in[2(N-1)/N,(N-1)]$. Очевидно, що найскладнішою у сенсі значення ϖ є структура, яка утворює повний граф, а найпростішою – послідовно поєднані вузли мережі. Загалом параметр ϖ дозволяє насамперед порівнювати складність моделей СМС однакової розмірності.

Співвідношення між розмірнісною та коннекційною складністю моделі мережі можна пояснити на наступному прикладі. Мережеві структури часто виникають в задачах обчислювальної математики. Порівняльний аналіз складності моделей цих структур дозволяє обирати найефективніші методи розв'язання конкретних прикладних проблем. Розглянемо задачу розрахунку електростатичного поля, яке генерується складним електрофізичним пристроєм та зводиться до розв'язання тривимірної граничної задачі для рівняння Лапласа [5]. Чисельний розв'язок цієї задачі можна шукати методами граничних елементів (МГЕ), скінченних різниць або скінченних елементів (МСЕ). Нехай область пошуку чисельного розв'язку — одиничний куб. Побудуємо у ньому та на граничній поверхні регулярну прямокутну сітку з кроком $h=10^{-2}$. Тоді у випадку застосування МГЕ необхідно розв'язати систему лінійних алгебраїчних рівнянь (СЛАР) зі щільною матрицею **G**, яку можна інтерпретувати як матрицю суміжності зваженої мережевої структури, вузлами якої при кусково-лінійній апроксимації шуканої густини потенціалу є вузли сітки на поверхні куба [6]. Розмірність N_G та кількість ненульових елементів L_G матриці **G** у цьому випадку визначаються рівностями $N_G \approx 6 \cdot 10^4$ та $L_G \approx 36 \cdot 10^8$. Застосування МСЕ при такому ж типі апроксимації шуканого розв'язку призводить до необхідності розв'язання СЛАР із семидіагональною матрицею **F** розмірності $N_F \approx 10^6$, яку можна інтерпретувати як матрицю суміжності зваженої мережевої структури, вузлами якої є вузли сітки в кубі. Кількість ненульових елементів цієї матриці $L_F \approx 7 \cdot 10^6$. Матриця **G** є матрицею суміжності зваженої мережі, яка утворює повний граф, а матриця **F** — зваженої мережі, у якій зв'язки існують лише між суміжними вузлами скінченно-елементної сітки в кубі. Користуючись термінами теорії складних мереж (ТСМ), елементи матриць **G** та **F** визначають силу взаємозв'язку між вузлами відповідних мережевих структур. Коннекційна складність моделей цих структур відповідно дорівнює $\varpi_G \approx 6 \cdot 10^4$ та $\varpi_F \approx 7$. Порівняльний аналіз розмірнісної та коннекційної складності цих моделей дає значення $N_G / N_F \approx 6 \cdot 10^{-2}$ та $\varpi_G / \varpi_F \approx 10^4$. Це означає, що складність структури сіткової області, отриманої за допомогою МГЕ, приблизно в 600 раз перевищує складність структури, отриманої за допомогою МСЕ. Розв'язання методом Гауса СЛАР із матрицею **G** потребує виконання приблизно у 20 разів більше арифметичних операцій порівняно з розв'язанням його аналогом СЛАР із матрицею **F**. Це пояснює значно більшу популярність для моделювання багатьох фізичних процесів саме методу скінченних елементів, оскільки складність мережевих структур, які вибудовуються під час його застосування є істотно меншою ніж для методу граничних елементів.

Редукція фіктивних елементів СМС. Серед багатомільйонної аудиторії користувачів соціальних мереж кількість активних не перевищує 10—30 % [7]. Активний словниковий запас пересічної англомовної людини не перевищує 10 тис. слів, хоча загальна їх кількість сягає півмільйона [8]. Google за запитом може видавати сотні тисяч або мільйони результатів пошуку, але пересічний користувач зазвичай обмежується переглядом кількох десятків або сотень цих результатів. Можна навести ще чимало прикладів реальних мереж, більшість вузлів та зв'язків яких насправді не беруть участі у функціонуванні системи. Водночас, досліджуючи систему та формуючи її модель, ми зацікавлені в чіткій ідентифікації структури СМС. Елементи, що включені до складу мережі, але не задіяні у процесі функціонування відповідної системи, називатимемо фіктивними.

Введемо потокову матрицю суміжності $V = \{V_{ij}\}_{i,\,j=1}^{N}$ системи, у якій вага $V_{ij} \in [0, 1]$ ребра, що поєднує вузли n_i та n_j, $i, j = \overline{1, N}$, дорівнює нормованому до 1 значенню об'ємів потоків, які проходять через згадане ребро за певний проміжок часу $[0, T]$. Визначити наявність фіктивних вузлів у СМС можна, розв'язавши задачу на власні значення для потокової матриці суміжності мережі. Цей підхід має два недоліки. А саме, розв'язання цієї задачі потребує виконання порядку $O(N^3)$ арифметичних операцій, що для загалом щільних матриць розмірності 10^6—10^{12} є практично невирішуваною проблемою. Цей підхід також не гарантує визначення усіх фіктивних зв'язків мережі. Потокова матриця суміжності дає можливість визначати у структурі вихідної СМС усі фіктивні елементи та видаляти їх з моделі структури. Як слідує із наведених вище прикладів, у багатьох випадках це дозволяє суттєво зменшити розмірність моделі системи без втрати її адекватності. Якщо певний елемент матриці суміжності **A** є ненульовим, але відповідний елемент матриці **V** дорівнює нулю, то це означає фіктивність відповідного ребра. Якщо серед усіх елементів i-го рядка та стовпця матриці суміжності мережі є ненульові, але всі елементи i-го рядка та стовпця матриці **V** є нульовими, то це означає фіктивність вузла n_i, $i = \overline{1, N}$. Зменшення розмір-



Фрагменти залізничної транспортної мережі західного регіону України, відповідної залізничної транспортної системи та їх структурної 4- та потокової 0,7-серцевин

нісної складності моделі СМС після вилучення фіктивних вузлів визначається параметром $\eta = \mathrm{rang}(A)/\mathrm{rang}(V)$, а складності за кількістю зв'язків — параметром $\mu = L_A / L_V$.

Спрощення моделі СМС без втрати її адекватності можна здійснювати не тільки шляхом виключення фіктивних вузлів та зв'язків, які входять до складу мережі, але й так званих вузлів-транзитерів. Найпростішим прикладом вузла-транзитера є вузол зі ступенем 2, у якому не відбувається відбір або додавання нових об'ємів потоків. Такими вузлами є проміжні станції у системі вантажних перевезень залізничної транспортної мережі, проксі-сервери в Інтернеті [9] тощо.

Потокові серцевини СМС. У ТСМ найбільш дієвим способом спрощення моделі бінарної мережі є використання поняття її k-серцевини, тобто найбільшої підмережі вихідної мережі, усі вузли якої мають ступінь не менший за k та вилучені зі структури мережі вузлів зі ступенем, меншим за k [10]. Потокова матриця суміжності СМС дозволяє ввести поняття потокової λ-серцевини мережевої системи [11], як найбільшої підмережі вихідної мережі, для якої усі елементи матриці **V** мають значення, не менші за λ, де $\lambda \in [0,1]$. На рисунку *а* схематично зображено структуру залізничної транспортної мережі західного регіону України з виключеними вузлами-транзитерами. Рисунок *б* містить той самий фрагмент з відображенням об'ємів руху потоків, величина яких є пропорційною до товщини ліній. На рисунку *в* та *г* зображені 4-серцевина цієї мережі та потокова 0,7-серцевина відповідної СМС. Структури цих серцевин істотно відрізняються, причому потокова λ-серцевина СМС дає значно важливішу інформацію для системних досліджень, ніж її k-серцевина. Так, відключення вузла *A* зі ступенем 3, який лежить на шляху інтенсивного руху потоків, призведе до значно більших проблем у функціонуванні залізничної системи, ніж відключення вузла *B* зі ступенем 4 і навіть вузла *C* зі ступенем 5. Розмірність моделей великих СМС можна зменшити, досліджуючи не всю мережу, а її λ-серцевину. Зображена на рисунку *а* залізнична транспортна мережа з виключеними вузлами-транзитерами містить 29 вузлів та 62 ребра (загалом до її складу входять 354 вузли та 421 ребро), а її 0,7-серцевина (рисунок *г*), яка забезпечує понад 70 % усіх перевезень, нараховує 4 вузли та 12 зв'язків. Тобто, зменшення складності моделі 4-серцевини порівняно з моделлю вихідної мережі визначається параметрами $\eta_{k=4} = 12{,}2$ та $\mu_{k=4} = 6{,}8$, а моделі потокової 0,7-серцевини — параметрами $\eta_{\lambda=0{,}7} = 88{,}5$ та $\mu_{\lambda=0{,}7} = 35{,}1$ відповідно.

Очевидно, що адекватність моделі потокової серцевини порівняно з моделлю всієї СМС є пропорційною її питомій вазі σ_λ в системі. Функціональний підхід до обчислення значення σ_λ полягає в наступному. Введемо потокову матрицю суміжності λ-серцевини за співвідношенням

$$V^\lambda = \{V_{ij}^\lambda\}_{i,j=1}^N, \quad V_{ij}^\lambda = \begin{cases} V_{ij}, & \text{якщо } V_{ij} \geqslant \lambda, \\ 0, & \text{якщо } V_{ij} < \lambda, \end{cases} \quad \lambda \in [0,1].$$

Параметр σ_λ визначає відношення об'ємів потоків, які проходять λ-серцевиною, до об'ємів потоків, які проходять мережею загалом за період $[0,T]$, а саме $\sigma_\lambda = s(\mathbf{V}^\lambda)/s(\mathbf{V})$, де $s(\mathbf{V})$ та $s(\mathbf{V}^\lambda)$ — сума усіх елементів матриць **V** та $\mathbf{V}^\lambda$ відповідно. Оскільки основною ціллю більшості мережевих систем є забезпечення руху певного типу потоків, то параметр σ_λ кількісно визначає наскільки λ-серцевина забезпечує реалізацію цієї цілі. Якщо замість моделі всієї системи ми досліджуємо модель її λ-серцевини, то значення σ_λ можна інтерпретувати як міру функціональної адекватності цієї моделі. У результаті виділення та дослідження потокової серцевини СМС приводить до аналізу або моделювання її функціонально найпріоритетніших підсистем. Так, поширення епідемій зазвичай відбувається на шляхах пересування великих мас людей, а розповсюдження комп'ютерних вірусів — на шляхах інтенсивного інформаційного трафіку. Моделі потокових серцевин таких СМС з великими значеннями λ визначають найбільш ймовірні шляхи розгортання таких процесів та дозволяють своєчасно їм протидіяти.

Інкапсуляція підмереж СМС. Серцево-судинна система людського організму включає в себе як основні, так і периферійні вени та артерії. Розрив однієї із основних судин може призвести до швидкого летального результату. У великому місті перекриття основних автомагістралей може призвести до колапсу всієї його транспортної системи. Аварії на магістральних лініях електропередач під час стихійних лих часто призводили до відключення від електроенергії цілих регіонів країни. Однак це зовсім не означає, що зв'язки з малим значенням λ можна повністю ігнорувати. Невеликі населені пункти, кількість яких є переважаючою у будь-якій країні, потрібно забезпечувати продуктами, транспортом, медичними, освітніми та іншими послугами, незважаючи на незначні об'єми потоків. Якщо довжина основних судин людського тіла сягає кількох кілометрів, то довжина його капілярної мережі є близькою до 100 тис. км. Однак саме капілярною мережею клітини людського тіла постачаються всіма необхідними для життєдіяльності організму речовинами та саме капілярами виводяться результати цієї життєдіяльності. Наведені приклади підводять нас до ще одного способу зменшення розмірності моделей структур СМС без істотної втрати адекватності, який полягає в інкапсуляції окремих її підмереж та їх заміщення так званими мета-вузлами. Для серцево-судинної системи людини це означає зменшення розмірності її моделі у тисячі разів. Інкапсуляцію доцільно використовувати до спільнот соціальної природи, інфраструктури населених пунктів під час дослідження автотранспортної системи регіону або країни тощо. Основні ознаки підмережі, яка може бути інкапсульованою, такі: 1) вона є зв'язною підмережею доповнення до відповідної λ-серцевини СМС; 2) кожний із елементів такої підмережі здійснює незначний вплив на систему, але вплив інкапсульованої області або заміщуючого її мета-вузла є співмірним із впливом інших елементів λ-серцевини СМС; 3) кількість зв'язків інкапсулюючого мета-вузла не перевищує максимального ступеня вузлів вихідної СМС. Очевидно, що інкапсуляція підмереж, які входять до доповнення до λ -серцевини СМС та включення мета-вузлів та їх зв'язків у структуру λ -серцевини дозволяє істотно збагатити модель цієї серцевини та посилити достовірність отриманих на її основі результатів дослідження системи.

Таким чином у статті проаналізовано проблему кількісної складності мережевих систем, яка постає під час побудови їх моделей. Запропоновано низку ефективних підходів до редукції складності моделей великих СМС з одночасним відстеженням збереження міри їх адекватності: вилучення фіктивних та транзитних елементів мережі, використання потокових серцевин та інкапсуляція їх доповнень. За допомогою введених в роботі показників розмірнісної та коннекційної складності показано, що пропоновані підходи дозволяють зменшувати кількісну складність моделей мережевих систем на порядки і більше.

ЦИТОВАНА ЛІТЕРАТУРА

1. Dorogovtsev S.N., Mendes J.F.F. Evolution of networks: From biological Nets to the Internet and WWW. Oxford: Oxford Univ. Press, 2013. 280 p.
2. Caldarelli G., Vespignani A. Large scale structure and dynamics of complex networks: From information technology to finance and natural science. New York: World Scientific, 2007. 251 p.
3. Spickermann C. Entropies of condensed phases and complex systems. Berlin: Springer, 2011, 223 p.
4. Суперкомпьютер несколько месяцев создавал модели спиральных галактик [Електронний ресурс]. Режим доступу: http://hyser.com.ua/tehnology/superkompyuter-neskolko-mesyatsev-sozdaval-modeli-spiralnyh-galaktik-200853.
5. Polishchuk O. Solution of double-sided boundary value problems for the Laplacian in R^3 by means of potential theory methods. *In Proc. of the XIX-th Intern. Seminar on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory*. 2014. P. 140-142.
6. Polishchuk O.D. On the removal of singularities in the numerical solution of integral equations of the potential theory. *J. Math. Sci.* 2016. № 1. P. 27-37. doi: https://doi.org/10.1007/s10958-015-2646-4
7. Prell C. Social Network Analysis: History, Theory and Methodology. New York: SAGE. 2012. 263 p.
8. Francis W.N., Kucera H. Frequency Analysis of English Usage. Boston: Houghton Mifflin, 1982. 213 p.
9. Polishchuk D., Polishchuk O., Yadzhak M. Complex evaluation of hierarchically-network systems. *Autom. Control and Inform. Sci.* 2014. **2** (2). P. 32—44. doi: https://doi.org/10.12691/acis-2-2-1
10. Dorogovtsev S.N., Goltsev A.V., Mendes J.F.F. k-core organization of complex networks. *Phys. rev. lett.* 2006. **96** (4). 040601. doi: https://doi.org/0.1103/PhysRevLett.96.040601
11. Поліщук О.Д., Яджак М.С. Мережеві структури та системи: II. Серцевини мереж та мультиплексів. *Систем. дослідження та інформ. технології.* 2018. № 3. С. 38—51. doi: https://doi.org/10.20535/SRIT.2308-8893.2018.3.04

Надійшло в редакцію 04.04.2019

REFERENCES

1. Dorogovtsev, S. N. & Mendes, J. F. F. (2013). Evolution of networks: From biological Nets to the Internet and WWW. Oxford: Oxford Univ. Press.
2. Caldarelli, G. & Vespignani, A. (2007). Large scale structure and dynamics of complex networks: From information technology to finance and natural science. New York: World Scientific.
3. Spickermann, C. (2011). Entropies of condensed phases and complex systems. Berlin: Springer.
4. The supercomputer created models of spiral galaxies for several months. Available: http://hyser. com.ua/tehnology/superkompyuter-neskolko-mesyatsev-sozdaval-modeli-spiralnyh-galaktik-200853 (in Russian).
5. Polishchuk, O. (2014). Solution of double-sided boundary value problems for the Laplacian in R^3 by means of potential theory methods. In Proc. of the XIX-th Intern. Seminar on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory, pp. 140-142.
6. Polishchuk, O. D. (2016). On the removal of singularities in the numerical solution of integral equations of the potential theory. J. Math. Sci. No. 1, pp. 27-37. doi: https://doi.org/10.1007/s10958-015-2646-4

7. Prell, C. (2012). Social Network Analysis: History, Theory and Methodology. New York: SAGE.
8. Francis, W. N. & Kucera, H. (1982). Frequency Analysis of English Usage. Boston: Houghton Mifflin.
9. Polishchuk, D., Polishchuk, O. & Yadzhak, M. (2014). Complex evaluation of hierarchically-network systems. Automatic Control and Information Sciences. Vol. 2 (2), pp. 32-44. doi: https://doi.org/10.12691/acis-2-2-1
10. Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. (2006). k-core organization of complex networks. Phys. Rev. Let., 96 (4), 040601. doi: https://doi.org/0.1103/PhysRevLett.96.040601
11. Polishchuk, O. D. & Yadzhak, M. S. (2018). Network structures and systems: II. Cores of networks and multiplexes. Systems research and information technologies. No. 3, pp. 38-51 (in Ukrainian). doi: https://doi.org/10.2053/SRIT.2308-8893.2018.3.04

Received 04.04.2019

А.Д. Полищук

Институт прикладных проблем механики и математики им. Я.С. Пидстрыгача НАН Украины, Львов
E-mail: od_polishchuk@ukr.net

РЕДУКЦИЯ СЛОЖНОСТИ МОДЕЛЕЙ СЕТЕВЫХ СТРУКТУР И СИСТЕМ

Анализируется проблема сложности сетевых структур и систем. Определяются количественные показатели сложности сети и их использование для выбора эффективной модели структуры системы. Предлагаются методы редукции сложности моделей систем с одновременным отслеживанием сохранения меры их адекватности. Эффективность предложенных подходов иллюстрируется на примерах реальных сложных систем.

Ключевые слова: *сетевая система, сложность, редукция, сердцевина, инкапсуляция, адекватность.*

O.D. Polishchuk

Pidstryhach Institute for Applied Problems of Mechanics and Mathematics of the NAS of Ukraine, Lviv
E-mail: od_polishchuk@ukr.net

REDUCTION OF THE COMPLEXITY OF MODELS
OF NETWORK STRUCTURES AND SYSTEMS

The problem of complexity of network structures and systems is analyzed. The quantitative indicators of the dimensional and connective network complexity are determined, and examples of their application for choosing an effective model of the system structure are given. The methods of reduction of the complexity of models of network systems are offered, taking into account that such systems can be investigated only in general. The first of these approaches consists in the identification and exclusion of fictitious elements from the network, i.e. nodes and edges are formally included in the structure, but not involved in the system operation. This allows us to reduce the complexity of many real system models by dozens. The concepts of flow adjacency matrix and the flow core of a network system determining the most functionally important components of it are introduced. In the simplest case, the flow cores allow us to exclude the transit nodes from the system model, i.e. elements which do not add or remove the part of flows that are moving through the network. The specific weight of the flow core determines how adequate is its model in comparison with the source network model. A number of examples show that the flow cores significantly reduce the complexity of system models. The method of encapsulation of the components of supplements to flow cores is proposed to increase the adequacy of their models. The main features of subnets that can be encapsulated are determined, and examples of real systems are given, for which the encapsulation method reduces the dimension of their models by dozens and more without significant loss of adequacy.

Keywords: *network system, complexity, reduction, core, encapsulation, adequacy.*