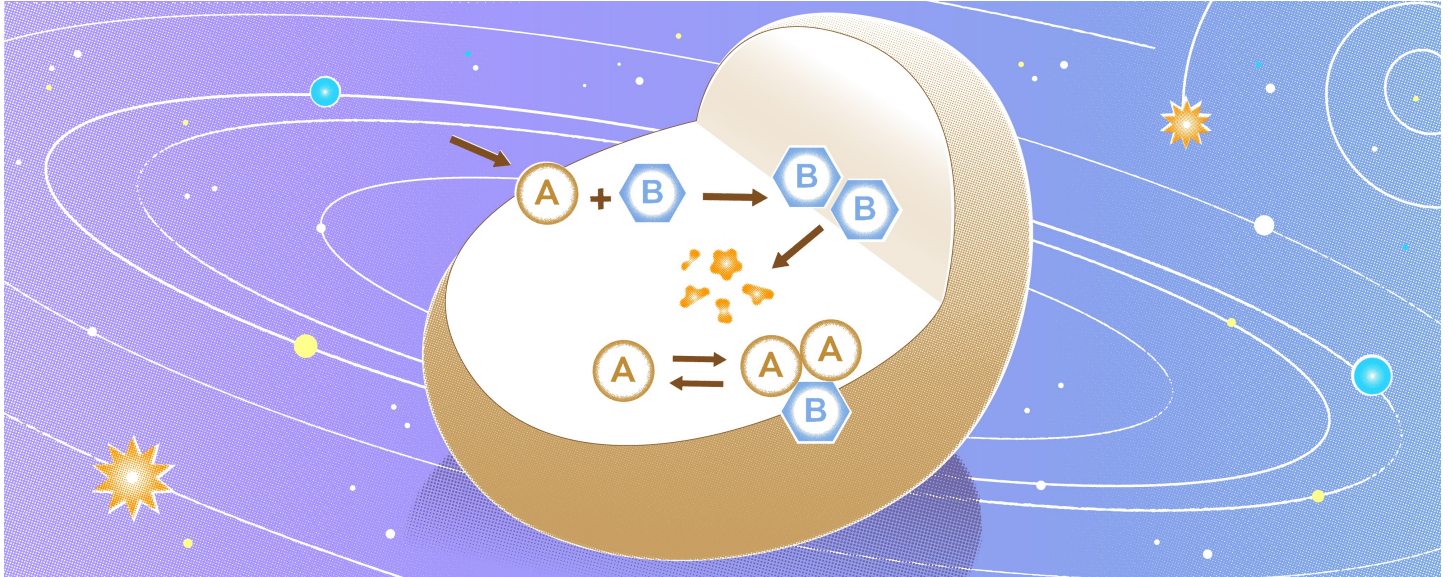


자연 현상을 표현하는 반응 네트워크

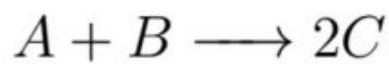
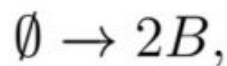
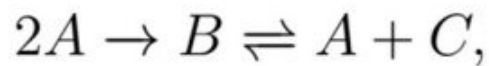
2022년 8월 26일

김진수



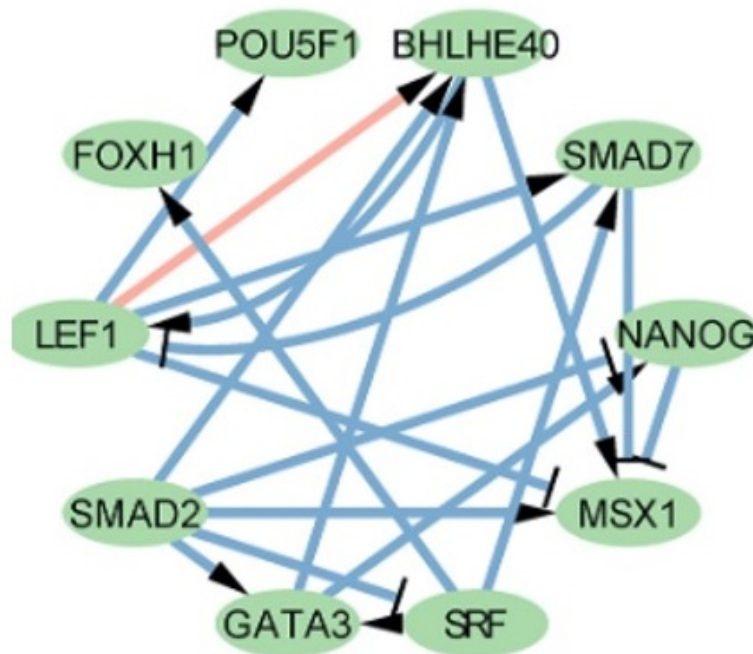
네트워크로 표현하는 자연 현상

우리가 학창 시절 화학 시간에 배운 $2H + O \rightarrow H_2O$ 와 같은 화학반응식을 그래프의 언어로 해석해보면 $2H + O$ 와 H_2O 를 마디점^{node}으로 그리고 그 둘 사이의 반응을(방향을 가지는) 변^{edge}으로 볼 수 있는데, 이런 반응을 한데 모아 그래프로 나타내면 화학반응식으로 구성된 네트워크가 형성된다. 이러한 네트워크를 화학반응네트워크 Chemical reaction network라고 한다. [그림1]에서 간단한 예를 소개한다.



[그림1]

화학반응식이라는 이름에서 오는 낯설음과는 달리 이 네트워크는 화학적 현상 뿐만 아니라 우리 주위에서 일어나는 많은 상호작용 혹은 자연현상을 표현할 수 있다. 그 예로, 최근 코로나19 사태를 분석하기 위해 사용된 SIR^{Susceptible, Infection, Recovered} 모델은 $S + I \rightarrow 2I, I \rightarrow R$ 로 나타낼 수 있다. 다른 예로 우리 몸속의 단백질이 생성되는 과정을 간단하게 표현한 DNA-mRNA-Protein 반응식도 $P \rightarrow \emptyset \rightleftharpoons mRNA \rightarrow mRNA + P$ 처럼 화학반응네트워크로 나타낼 수 있다 (여기서 \emptyset 는 주어진 시스템의 외부를 의미한다. 즉 $P \rightarrow \emptyset$ 는 P 가 사라지는 반응을 의미한다. 마찬가지로 $\emptyset \rightarrow A$ 라는 반응은 A 의 생성으로 볼 수 있다.). 이러한 이유로 화학반응네트워크의 더 일반적인 개념을 위해 생화학반응네트워크^{Biochemical reaction networks}, 상호작용 네트워크, 혹은 반응네트워크 등을 사용하기도 한다. 이 글에서는 간략하게 반응네트워크로 부르기로 하겠다.



[그림2]

수학 뿐만 아니라 기타 과학, 공학 분야에서 네트워크를 기반으로 하는 연구가 많이 이루어지고 있는데 다른 네트워크 모델들과 달리 반응네트워크는 반응식을 통해 실제 변수들의 상호작용을 상세히 표현할 수 있다. [그림2]¹의 네트워크는 유전자와 단백질 간의 관계를 나타내는 유전자 조절 네트워크^{Gene-regulatory network}인데 여기서는 각 마디점 (유전자 혹은 단백질)들이 서로 상관 관계가 있으면 변으로 연결시키고 그 상호작용의 특성 (활성화 혹은 억제 작용)에 따라 화살표 끝의 모양(일반적인 화살표이나 끝이 닫힌, 즉 T자를 눕혀 놓은 것과 같은 모양)을 다르게 한다. 하지만 이에 대응하는 반응네트워크는 상관 관계 유무와 더불어 그 개체들 간의 반응 $A + B \rightarrow C$ 와 같은 실제 상호작용을 표현한다.

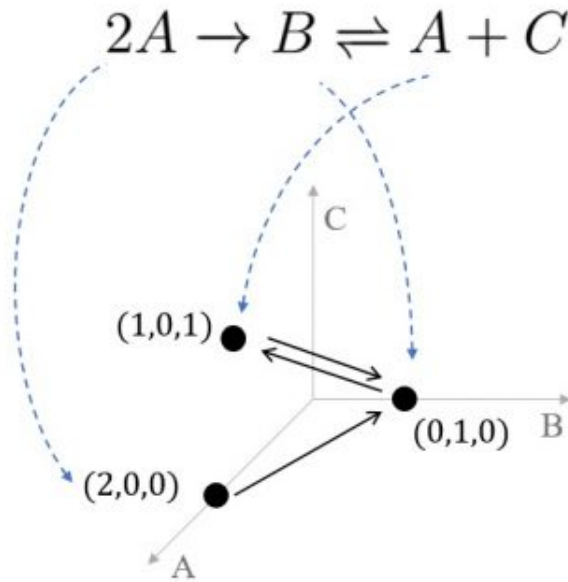
위에서 소개한 예시에서 추측할 수 있듯이, 반응네트워크는 생명, 생태계 관련 시스템을 수학적 모델을 통해 분석하는

¹Zheng, Xiao, Yuan Huang, and Xiufen Zou. "scPADGRN: A preconditioned ADMM approach for reconstructing dynamic gene regulatory network using single-cell RNA sequencing data." PLoS computational biology 16.7 (2020): e1007471.

수리생물학 연구에서 자주 쓰이는 도구이다. 추측하자면 독자분들께서 이미 접해보신 수리생물학의 연구 방법은 아마 특정 생화학시스템 모델 하나 혹은 몇 개를 정해두고 그 시스템들을 묘사하는 수학모델을 만든 후 그것을 미분방정식, 확률모델, 통계모델, 컴퓨터 시뮬레이션 등을 이용해 분석하는 접근일 것이다. 이러한 접근법과는 조금 다르게 반응네트워크를 수학적으로 연구하는 화학반응 네트워크 이론^{Chemical reaction network theory} 분야 연구자들은 수학적으로 표현된 정성적인 현상들(해의 존재성, 유일성, 안정성 등)을 어떤 생물현상이 가질 수 있는지 역으로 생각해보는 새로운 패러다임을 수리생물학에 제시한다. 이번 연재에서는 이 새로운 패러다임에 대한 설명과 그것의 예시로서 한 가지 중요한 이론을 소개하겠다.

반응네트워크의 용어와 결정론적 모델링

먼저 반응네트워크에 관련된 기본적인 용어들을 소개하겠다. 반응네트워크의 마디점들은 $2A + B$ 처럼 변수들에 양의 정수를 곱해 더하는 선형결합으로 나타나는데 이때 각 변수들을 종 혹은 개체^{species}라 부르고 그들의 선형결합을 복합체^{complex}라 부른다. 두 복합체들 간의 반응은 말 그대로 반응^{reaction}이라 부른다. 즉 반응은 $2A + B \rightarrow C$ 처럼 한 복합체가 다른 복합체로 변하는 과정을 통해 어떤 개체는 사라지고 어떤 개체는 생성되는 것을 표현하는 것이다. 화학반응에서 몇 개의 개체가 합성되고 생성되는지를 나타내는 수치를 화학량^{stoichiometry}이라 하는데 이는 간단하게 양의 정수로 이루어진 벡터들로 표현된다. 가령 앞서 예로 들었던 세 개의 개체로 이루어진 반응 $2A + B \rightarrow C$ 에서 원료가 되는 복합체^{source complex} 혹은 reactant $2A + B$ 를 벡터 $y = (2, 1, 0)^\top$, 결과물 복합체^{product complex} C 는 벡터 $y' = (0, 0, 1)^\top$ 로 간략하게 나타낼 수 있다. 마지막으로 벡터 $y' - y = (-2, -1, 1)^\top$ 는 2개의 A 와 1개의 B 를 잃고 1개의 C 를 얻는 반응 자체를 표현하는데 이러한 벡터를 반응벡터 (혹은 화학량 벡터^{stoichiometry vector})라 부른다. 앞서 소개한 이 3가지 벡터로 화학량에 관한 정보를 모두 제공한다. 화학량은 반응네트워크를 유클리드 공간에 어떻게 매장^{embed}해야 하는지도 알려주기도 한다[그림3].



[그림3]

김진수

반응네트워크로 표현된 생화학현상을 여러 수학모델을 이용해 분석할 수 있는데 흔히 결정론적 방법과 확률론적 방법으로 분류된다. 이번 연재에서는 결정론적 방법에 해당하는 상미분방정식 모델을 주로 다뤄보겠다. 먼저 실제 생화학시스템이 존재하는 물리적인 공간 (실험 용기, 세포 핵 등)이 잘 섞여 있다고 가정하면 공간의 각 지점에서 일어나는 반응의 크기, 개체들의 농도가 모두 일정하므로, 이 시스템을 분석하기 위해서는 공간에 관한 변화 없이 시간에 관한 변화만 따지면 된다. 이때 반응네트워크에 나타나는 반응들이 시간이 흐름에 따라 연속적으로 발생한다고 가정하면 해당 시스템에 속한 개체들 ([그림3]에서 A 나 B 와 같은 변수들)의 농도도 그에 따라 연속적으로 변한다. 그리하여 우리는 이러한 시간에 따른 농도의 변화를 상미분방정식으로 표현할 수 있다. 예를 들어 간단한 반응네트워크 $2A \xrightleftharpoons[2]{} B$ 를 생각해보자. 위에서 이야기했던 가정들을 전제하면, A 와 B 의 농도를 각각 나타내는 벡터 $x(t) = (x_A(t), x_B(t))$ 는 아래와 같은 미분방정식을 만족시키는 것이 자연스럽다.

$$\text{Eq1: } \frac{d}{dt}x(t) = \mathcal{K}_{2A \rightarrow B}(x) \begin{pmatrix} -2 \\ 1 \end{pmatrix} + \mathcal{K}_{B \rightarrow 2A}(x) \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

우변의 첫 번째 항은 반응 $2A \rightarrow B$ 의 화학량 벡터 $(-2, 1)^\top$ 에 음이 아닌 함수 $\mathcal{K}_{2A \rightarrow B}(x)$ 가 곱해진 형태이다. 이 항이 의미하는 것은 $x(t)$ 가 2개의 A 를 잃고 하나의 B 를 얻는 방향으로 국소적으로 끌려가는데 그 크기 (혹은 속력)가 $\mathcal{K}_{2A \rightarrow B}(x)$ 인 것이다. 마찬가지로 두 번째 항은 $x(t)$ 가 반응 $B \rightarrow 2A$ 의 화학량 벡터 $(2, -1)^\top$ 에 해당하는 방향으로 속도 $\mathcal{K}_{B \rightarrow 2A}(x)$ 로 끌려간다는 뜻이다. 반응의 크기를 결정하는 $\mathcal{K}_{2A \rightarrow B}(x)$ 나 $\mathcal{K}_{B \rightarrow 2A}(x)$ 와 같은 반응 속도 함수 reaction rate function 들은 주로 질량 작용 역학 mass action kinetic²이나 hill 함수 등을 주로 선택한다.

²Mass action kinetics에 따르면 $\mathcal{K}_{2A \rightarrow B}(x) = x_A^2$ 그리고 $\mathcal{K}_{B \rightarrow 2A}(x) = x_B$ 이다.

³미분방정식 $\frac{d}{dt}x(t) = f(x(t))$ 에서 $f(x^*) = 0$ 을 만족시키며 모든 좌표가 양수인 점 x^* .

$2A \xrightleftharpoons[2]{3} B$ 에 등장한 2나 3과 같은 상수가 반응 함수에 포함되는데 이러한 상수들을 반응 속도 상수 reaction rate constant 라 한다. 이것을 일반화하면 $x_i(t)$ 를 주어진 반응네트워크의 i 번째 개체의 농도라고 한다면 $x = (x_1(t), x_2(t), \dots, x_d(t))^T$ 는 미분방정식

Eq2:

$$\frac{d}{dt}x(t) = \sum_{i=1}^r \kappa_i \mathcal{K}_i(x) \eta_i.$$

만족시킨다. 여기서 $\kappa_i, \mathcal{K}_i(x), \eta_i$ 는 각각 i 번째 반응의 반응 속도 상수, 반응 속도 함수 그리고 반응 벡터이다.

반응네트워크 이론의 철학

아마도 어떤 독자분들은 위에서 소개된 상미분방정식 모델을 보고 어떤 궁금증이 생겼을 수도 있다. 예상하자면, “생화학시스템을 곧바로 미분방정식으로 표현하지 않고 왜 굳이 반응네트워크로 표현하고 그것의 화학량 벡터들을 따지는 (혹은 유클리드공간에 매장시키는) 번거로운 과정을 거치는가?” 라는 질문일 것이다. 이 질문에 대한 한 가지 답은 바로 어떤 경우에는 미분방정식을 직접 분석하지 않고 반응네트워크의 구조만으로 미분방정식의 해가 가지는 안정성, 정상 상태 steady states의 존재성 등을 알아낼 수 있기 때문이다. 여기서 네트워크의 구조는 가령 복합체들의 개수, 연결체들의 개수, 특정 반응벡터의 유무와 같이 주어진 시스템이 가지는 생화학적 의미가 아닌 네트워크를 유클리드공간에 매장한 후 그것의 ‘겉모양’으로만 결정되는 요소들을 의미한다. 이를 구체적으로 설명하기 위해 반응네트워크 분야에서 가장 역사적으로 중요한 무결핍 Deficiency zero 이론을 소개하겠다.

“주어진 생화학시스템을 반응네트워크로 표현했을 때, 그 네트워크의 구조가

$$\text{Deficiency} = (\text{마디점의 개수}) - (\text{연결체의 개수}) - (\text{반응벡터들로 만든 행렬의 rank}) = 0$$

를 만족한다고 가정하자. 그러면 주어진 반응 상수들과는 무관하게 해당 시스템 표현하는 질량 작용 역학을 따르는 상미분방정식의 양의 정상상태³는 유일⁴하고 국소적으로 안정적⁵이다.”⁶

⁴정확하게 말하면, 화학량적 호환 등급 stoichiometry compatibility class이라는 영역에서 유일하다. 하지만 이 영역은 개체들간의 보존 관계 (가령 개체 A 와 B 의 농도의 합이 늘 일정한 관계)가 있는 특별한 경우를 제외하고는 늘 양수 유클리드 공간 전체가 된다. 자세한 사항은 각주 6에 소개된 논문을 참고하길 바란다.

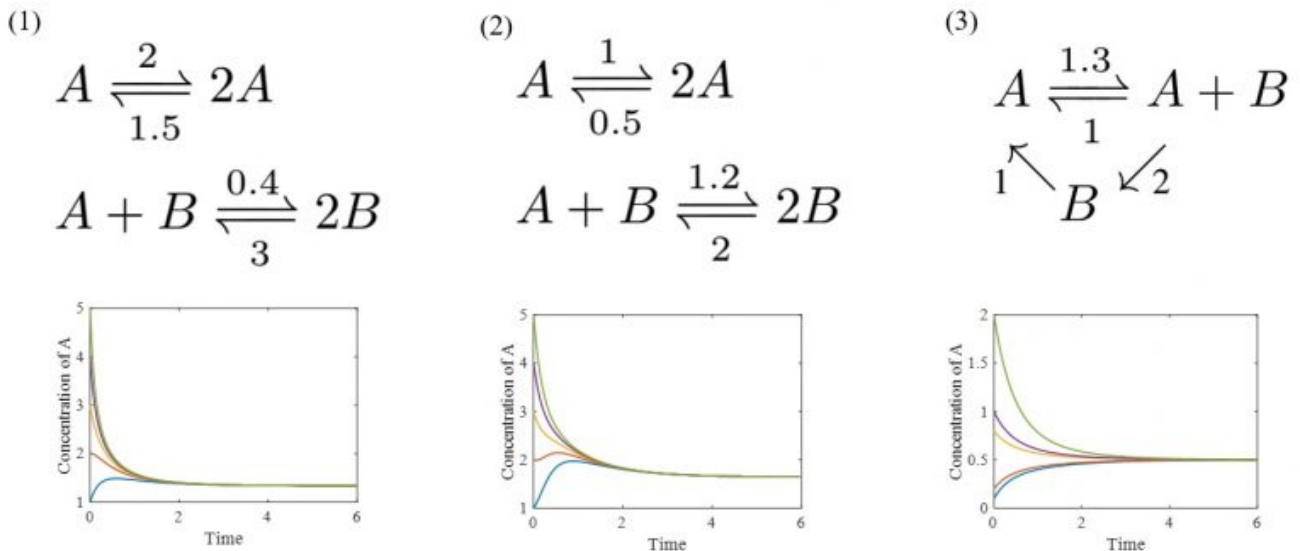
⁵ 정상상태 x^* 근처에서 출발한 미분방정식의 해가 시간에 따라 x^* 로 수렴함.

⁶ M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors. I. The deficiency zero and deficiency one theorems, Chem. Eng. Sci., 42 (1987), pp. 2229–2268.

위 이론에서 등장하는 몇 가지 생소한 개념들은 모두 후반부에 자세히 설명을 하겠다. 지금은 일단 위 이론에 숨어있는 철학을 파악하는 것에 목적을 두자. 위의 이론에서 특이한 점은 미분방정식의 정상 상태^{steady state}의 존재성 유일성 그리고 안정성이라는 중요한 특성이 주어진 반응 상수들에 관계없이 네트워크 구조로만 결정되었다는 점이다.

이와 같은 구조적 특성을 이용한 시스템의 분석은 생화학시스템을 개별적으로 연구하는 것이 아닌 특정 동역학적^{dynamical} 성질을 야기시키는 반응네트워크의 구조를 찾음으로써 그 구조를 가진 모든 생화학시스템을 하나의 집단으로 묶을 수 있게 한다. 그리고 이러한 접근법이 다른 수리생물학의 연구와 차별화되는 반응네트워크 이론의 철학 중 하나이다. 또한 “구조적 특성=반응 상수와 무관한 성질”로 생각할 수 있으므로 생화학시스템의 보편성^{universality}(동역학 시스템에서 매개변수와 관계없이 늘 일어나는 현상 따위)을 연구하는 것으로도 연결된다.

이 사실을 좀 더 구체적으로 설명하기 위하여 세 가지 반응네트워크를 예시로 들어보겠다. [그림4]에서 세 개의 반응네트워크는 각각 다른 화학 시스템을 표현한다.



[그림 4]

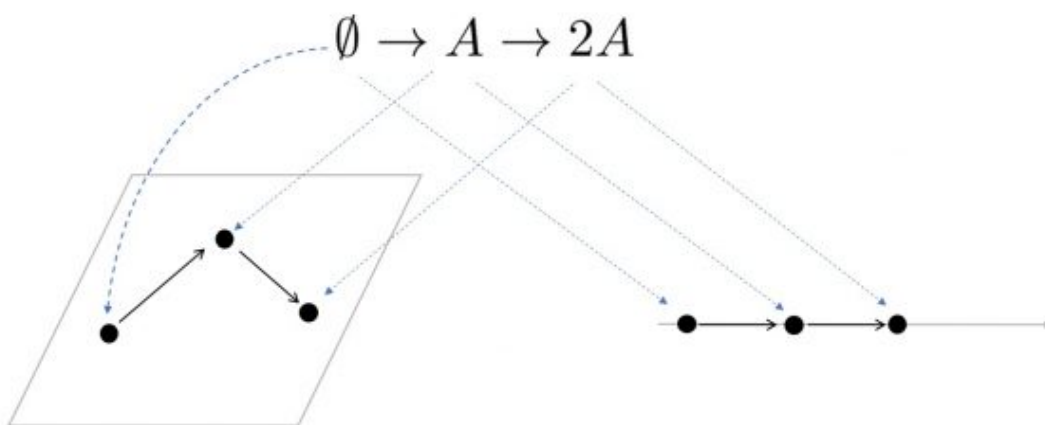
김진수

1번과 2번 반응네트워크는 네트워크의 구조는 같지만 반응 상수들이 다르고, 3번 네트워크는 앞의 두 네트워크들과 확연히 다르게 보인다. 먼저 1번과 2번 네트워크를 유클리드 공간에 매장하면 정확히 같은 네트워크를 얻는다. 그러므로 위에서 소개한 이론과 같이 구조적 특성으로만 결정되는 현상들은 1번과 2번이 모두 공통적으로 가지게 된다. 1번과 3

번 네트워크는 서로 다른 구조를 가졌지만 하나의 공통점을 가진다. 바로 위의 이론이 요구하는 조건 $\text{deficiency}=0$ 을 만족시킨다는 점이다. 결론적으로 위 세 가지 네트워크는 매개변수나 네트워크 구조에서 상이한 점이 있지만 $\text{deficiency}=0$ 이라는 구조적 특성을 모두 만족하므로, 각 모델을 표현하는 상미분방정식의 정상 상태는 유일성과 국소적 안정성이라는 공통점을 가진다. 이것을 확장하면 $\text{deficiency}=0$ 이라는 구조적 특성을 만족시키는 모든 생화학시스템을 하나의 집단으로 묶을 수 있고 그 집단에 속한 시스템은 양의 정상 상태의 유일성, 국소적 안정성이라는 공통적인 현상, 즉 보편성(universality)을 가지는 것이다. [그림4]의 그래프는 각 모델의 개체 A가 특정 범위의 초기값 (위의 그래프에서는 5개의 초기값을 임의로 선택함)을 가지면 언제나 유일한 양의 정상 상태로 수렴한다는 사실을 보여준다.

무결핍 이론

이제 위에서 언급한 무결핍 이론을 좀 더 살펴보자. 결핍(Deficiency)의 정의 (마디점의 개수)-(연결체의 개수)-(반응 벡터로 이루어진 행렬의 rank)에서 먼저 연결체란 네트워크 상에서 반응들로 연결된 마디점들의 집합을 의미한다. 예를 들면, [그림3]에 있는 1번 반응 네트워크에서 (방향을 잠시 잊고) 서로 연결된 마디점들을 다 모으면 총 2개의 집합 ($\{A, 2A\}$ 그리고 $\{A + B, 2B\}$), 즉 2개의 연결체가 있는 것이다. 이제 반응 벡터로 이루어진 행렬은 네트워크를 유클리드 공간에 매장 한 후 거기서 나오는 반응 벡터들을 열로 가지는 행렬을 의미한다. 이 행렬의 rank는 행벡터 (혹은 열벡터)들 중 서로 선형적으로 독립적인 벡터들의 수와 같은데, 이것은 반응 벡터들을 이용해 생성(span)된 벡터 공간⁷의 차원과도 같다. 결핍을 직관적으로 이해하는 방법 중 하나는 그것이 차원의 결핍을 의미한다고 생각하는 것이다.



[그림5]

김진수

예를 들어 [그림5]처럼 반응네트워크 $\emptyset \rightarrow A \rightarrow 2A$ 를 생각해보자. 이 네트워크를 (어떠한 화학적 해석이 없는) 추상적인 그래프로 보면, 이것은 2차 평면에 자연스럽게 매장할 수 있는 그래프가 된다. 하지만 주어진 화학량으로 인해 복합체 $\emptyset, A, 2A$ 는 1차원 유클리드 공간속에 각각 0, 1, 2에 대응되면 이 반응네트워크는 최종적으로 직선의 형태로 매장

되어야 한다. 그러므로 반응 네트워크를 화학적 대상으로 바라보면서 그것의 차원을 하나 잃게 되는데, 이것이 왜 이 반응 네트워크의 결핍이 1⁸인지를 설명해준다.

⁷ 반응 벡터 $\eta_1, \eta_2, \dots, \eta_r$ 들로 만들어진 선형 결합 $\sum_{i=1}^r a_i \eta_i$ 를 모두 모아 놓은 공간.

⁸ 마디점의 개수=3, 연결체의 개수=1, 반응 벡터로 이루어진 행렬 [1 1]의 rank=1.

즉 결핍은 반응네트워크를 '자유도'가 가장 높은 추상적인 그래프로 생각했을 때 그래프가 표현하는 차원과 화학적 해석이 결합했을 때의 차원간의 차이로 해석할 수 있다. 만일 $\emptyset \rightsquigarrow A$ 과 같이 결핍이 0이 되는 경우는 바로 그 차원의 차이가 0이라서 화학적 해석을 고려하더라도 반응네트워크가 가장 높은 자유도를 가지고 있다는 뜻이다. 그리고 그런 경우 무결핍이론이 설명하듯 개체들의 정상 상태의 유일성과 안정성이 보장된다.

무결핍의 의미는 설명하였지만 아마 여전히 독자들은 이 무결핍이 어떻게 실제로 상미분방정식의 정상 상태의 유일성과 안정성과 연관 되는지 상상하기 어려울 듯하다. 자세한 내용은 Feinberg 1987을 참조해야 하지만, 간단하게 이야기하면 아래와 같은

Eq3:

$$\frac{d}{dt}x(t) = \sum_{i=1}^r \kappa_i \mathcal{K}_i(x) \eta_i = Y A_\kappa \psi(x).$$

상미분방정식을 두 개의 행렬 Y 와 A_κ , 그리고 하나의 비선형 함수 $\psi(x)$ 로 나눈 후 이 두 개의 행렬의 대수적인 구조를 보면 바로 $\dim(Ker Y \cap Im A_\kappa) \leq$ 결핍 이라는 부등식이 나온다⁹, 이때 결핍이 0이 되어 공간 $Ker Y \cap Im A_\kappa$ 의 차원이 0이 되면 모든 정상 상태가 복합체 균형^{complex balance} (각 복합체로 들어오는 흐름과 나가는 흐름의 힘이 균형을 이루는 상태)라는 조건을 만족시키고 이 균형은 안정성과 유일성을 보장한다.

⁹ M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors. I. The deficiency zero and deficiency one theorems, Chem. Eng. Sci., 42 (1987).

이번 연재에서는 수리생물학의 도구 중 하나인 반응네트워크 속 개체들의 농도를 상미분방정식을 이용해 분석할 때 네트워크의 구조만을 가지고 어떤 의미 있는 정성적인 움직임을 찾을 수 있는지를 알아봤고 그 예시로 무결핍 이론을 이용했다. 그리고 이 같은 특이한 방향성, 즉 매개변수와 무관한 구조적인 특성만 가지고 생화학현상을 분석하는 접근법은 생화학시스템의 일반성을 찾는 것 과도 밀접한 연관이 있다는 것을 강조하였다. 다음 연재에서는 무결핍 이론 외에 다른 구조적특성과 시스템의 정성적인 움직임과의 관계를 소개하겠다. 또 확률 모델을 이용해 반응네트워크를 분석하는 방법을 이야기해보겠다.